

PRÉDICTION SPATIALE EN PRÉSENCE D'ERREURS SUBSTITUTIVES ET MODÉLISATION PAR DRAP STABLE

THÈSE N° 3758 (2007)

PRÉSENTÉE LE 20 AVRIL 2007

À LA FACULTÉ DES SCIENCES DE BASE

Chaire de statistique appliquée

SECTION DE MATHÉMATIQUES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Baptiste FOURNIER

ingénieur mathématicien diplômé EPF
de nationalité suisse et originaire de Nendaz (VS)

acceptée sur proposition du jury:

Prof. T. Liebling, président du jury
Prof. S. Morgenthaler, directeur de thèse
Prof. A. C. Davison, rapporteur
Prof. M. Genton, rapporteur
Dr A. Papritz, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Lausanne, EPFL

2007

*Tiger's got to hunt
Bird's got to fly
Man's got to sit
and wonder "why, why, why?"*

*Tiger's got to sleep
Bird's got to land
Man's got to tell himself
he understands*

Kurt Vonnegut Junior

Remerciements

J'aimerais tout d'abord exprimer mes remerciements au Professeur Stephan Morgenthaler pour m'avoir offert la possibilité d'accomplir cette thèse, pour sa confiance, son soutien et ses conseils accordés tout au long de ce travail.

Je remercie le Professeur Anthony Davison (EPFL), le Professeur Marc Genton (Université de Genève) et le Docteur Andreas Papritz (ETHZ) pour avoir accepté de constituer mon jury et avoir apporter un regard critique à mon travail. Je remercie également le Professeur Thomas Liebling (EPFL) d'avoir présider ce jury.

Je souhaite également remercier toutes les personnes qui d'une manière ou d'une autre m'ont accompagné au cours de ces années. Tout particulièrement, Grégoire Aubry pour son écoute et son support, Thomas Gsponer pour les conseils, les discussions et la relecture de ce document, Pascal Gillioz, pour toutes les bouffées d'oxygène prises ensemble en montagne et bien entendu mes parents pour m'avoir permis d'arriver jusque là.

Finalement, un énorme MERCI à Valérie pour sa présence, son soutien inconditionnel, sa patience, ses encouragements et ... le plus beau cadeau du monde !

Abstract

In geostatistics, the presence of outlying data is more the rule than the exception. Moreover, the statistical analysis of data contaminated by outliers requires caution, particularly when a spatial dependence exists. In order to take into account these possible outliers during the adjustment of the spatial process, a new modeling tool, called the substitutive errors model, is proposed. The optimal prediction in the least squares sense is derived and its properties are studied. Because of its complexity, this estimator needs in practice to be numerically approximated. An automated algorithm is proposed in this thesis. This method is based on an ordering of the observations with respect to the specified spatial process of interest, with the values least in agreement being included towards the end of the ordering. It proves to be useful in case of masked multiple outliers or nonstationary clusters. Simulations are carried out to illustrate its performances and to compare it to other forecasts, robust or not. An application to real data is provided as an illustration of its practical usefulness.

The second part of this work also deals with the presence of spatial heterogeneity. One could say that the proposed model offers a characterization of this heterogeneity rather than estimating the locations and sizes of outliers. It is based on the theory of bidimensional α -stable motion. This represents a generalization of the unidimensional brownian motion. In particular, the stability parameter α can be seen as a measure of the distance between the observations and the hypothesis of a Gaussian distribution. A method of estimation for the parameters of such a process is presented, based on a numerical constrained optimization of the likelihood. Its performances are illustrated by means of simulations. An application ends this second part.

Keywords : bidimensional α -stable motion, forward search, geostatistics, heterogeneity, robust kriging, substitutive errors.

Résumé

L'existence de données aberrantes en géostatistique tient plus de la règle que de l'exception. Le traitement statistique de valeurs isolées nécessite la plus grande prudence, spécialement en présence d'une dépendance spatiale. Afin de tenir compte de ces éventuels présences dans l'ajustement du processus spatial sous étude, une nouvelle modélisation, dite à erreurs substitutives, est proposée. L'estimateur optimal au sens des moindres carrés est dérivé et ses propriétés étudiées. En raison de sa complexité, cet estimateur doit en pratique être approché. Une méthode d'approximation entièrement automatisée est proposée. Elle se base sur un ordonnancement des données, de la plus en accord avec le processus intérêt à la plus éloignée. Cette méthode a l'avantage de bien se comporter même en cas d'effet de masquage ou en cas de valeurs aberrantes regroupées. Des simulations illustrent les performances de l'estimateur et le comparent à d'autres estimateurs de krigeage, robuste ou non. Une application à des données réelles est proposée pour illustration.

La seconde partie de ce travail traite également de la présence d'hétérogénéité spatiale. Plutôt que fournir un traitement, une caractérisation de cette hétérogénéité est proposée au travers d'une nouvelle modélisation. Cette dernière se base sur la théorie des mouvements α -stables bidimensionnels. Ils représentent une généralisation à deux dimensions du mouvement brownien. En particulier, le paramètre α peut être vu comme une mesure de l'éloignement des données de l'hypothèse d'une distribution gaussienne. Une méthode d'estimation pour les paramètres d'un tel mouvement est proposée. Basée sur une optimisation numérique sous contrainte de la vraisemblance, ses performances sont éprouvées au travers de simulations. Une application clos cette seconde partie.

Mots-clés : erreurs substitutives, géostatistique, hétérogénéité, méthode *forward search*, mouvement α -stable bidimensionnel, krigeage robuste.

Table des matières

Remerciements	i
Abstract	iii
Résumé	v
Table des figures et liste des tableaux	xii
Introduction	1
1 Géostatistique	5
1.1 Processus univarié	6
1.1.1 Stationnarité	7
1.1.2 Ergodicité	8
1.1.3 Variogramme	9
1.1.4 Anisotropie	12
1.2 Estimation et ajustement de la variation spatiale	13
1.2.1 Estimation	13
1.2.2 Ajustement	15
1.3 Krigage	17
1.3.1 Krigage robuste	20

2	Modèle à erreurs substitutives	23
2.1	Variogramme associé au modèle	26
2.2	Contamination gaussienne	27
2.2.1	Le modèle	29
2.2.2	Etape 1 : Détermination d'un scénario de contamination initial . . .	34
2.2.3	Etape 2 : Détermination d'un sous-ensemble de scénarios	39
2.2.4	Propriétés de l'estimateur	41
2.2.5	Simulations	45
2.3	Contamination non gaussienne	53
2.3.1	Distributions elliptiques	54
2.3.2	Le modèle	56
2.3.3	Simulations	59
2.4	Application	61
2.5	Dépendance spatiale dans les scénarios	66
2.5.1	Simulations	70
3	Drap stable	75
3.1	Distributions infiniment divisibles	75
3.2	Théorème Central Limite généralisé et domaines d'attraction	78
3.3	Les distributions stables	81
3.3.1	Description générale	81
3.3.2	Fonction caractéristique	82
3.3.3	Propriétés d'une loi stable	82
3.3.4	Fonctions de densité	85
3.3.5	Simulation	86
3.3.6	Estimation	87
3.4	Drap stable	89
3.4.1	Le drap brownien	91
3.5	Ajustement d'un drap stable	95
3.5.1	Extraction d'un échantillon	96
3.5.2	Optimisation sous contrainte de la log-vraisemblance	96
3.6	Simulations	103
3.7	Application	105

Conclusion	115
Bibliographie	115
Curriculum Vitae	121

Table des figures

1.1	Représentations de différents variogrammes isotropiques.	10
1.2	Processus intrinséquement stationnaires en fonction de l'étendue.	12
1.3	Représentation des poids optimaux du krigeage ordinaire.	18
1.4	Surfaces ajustées par krigeage ordinaire.	19
1.5	Surfaces ajustées par krigeage robuste.	21
2.1	Illustration unidimensionnelle de la prédiction.	29
2.2	Réalisation d'un modèle à erreurs substitutives.	34
2.3	Comparaisons des scénarios retenus.	35
2.4	Ajustement du variogramme de manière robuste.	36
2.5	Carrés des résidus de prédiction standardisés.	37
2.6	Evolution temporelle des résidus standardisés associés aux différents sites. .	38
2.7	Détermination du scénario initial fourni par <i>forward search</i>	38
2.8	Poids associés à l'ensemble des sommets de l'hypercube.	40
2.9	Comparaison des transformations probit des distributions empiriques des quatre estimateurs I.	48
2.10	Présentation de la disposition des sites pour les simulations.	50
2.11	Comparaison des surfaces de variance de prédiction.	51
2.12	Comparaison des sommes moyennes des résidus de prédiction au carré. . .	52
2.13	Représentation unidimensionnelle des variances de prédictions.	53
2.14	Représentation unidimensionnelle des résidus carrés de prédiction	53
2.15	Histogramme du pourcentage de scénarios retenus.	54
2.16	Comparaison des transformations probit des distributions empiriques des quatre estimateurs II.	60
2.17	Visualisation tridimensionnelle des surfaces ajustées pour <i>SIC04</i>	64
2.18	Comparaison de toutes les méthodes ayant participé à la compétition <i>SIC04</i>	65

2.19	Surface de prédiction obtenue par moyenne simple sur les plus proches voisins.	66
2.20	Représentation unidimensionnelle des prévisions	67
2.21	Configurations de contaminations envisagées.	70
2.22	Comparaison des transformations probit des distributions empiriques des quatre estimateurs III.	71
2.23	Représentation planaire des différentes surfaces obtenues pour <i>SIC04</i>	73
3.1	Représentation de densités stables avec $\beta = 0$	81
3.2	Représentation de densités stables avec $\beta = 0.5$	83
3.3	Représentation de densités stables en fonction de α et β	84
3.4	Représentation des cas particuliers parmi les densités stables	86
3.5	Différentes réalisations de draps stables.	95
3.6	Fonction de relaxation des contraintes concernant l'indice de stabilité . . .	98
3.7	Simulation des estimateurs donnés par (3.15).	100
3.8	Diagrammes en boîte des estimations obtenues; cas $\mu_0 = 0$, et $\sigma = 1$	104
3.9	Diagrammes en boîte des estimations obtenues; cas $\mu_0 = 3$, et $\sigma = 4$	107
3.10	Surface de rendement de blé en livre et surface des résidus	107
3.11	Ajustement du variogramme pour les données de Mercer and Hall (1911) .	108
3.12	Traitement géostatistique des données de Mercer and Hall (1911)	109
3.13	Logarithme du rapport de vraisemblance profilée pour α	110
3.14	Diagrammes des quantiles des résidus de prédiction standardisés.	111

Liste des tableaux

2.1	Distribution bivariée du processus de localisation de la contamination $B(\cdot)$.	27
2.2	Erreur absolue moyenne dans le cas d'une contamination gaussienne. . . .	47
2.3	Racine de l'erreur carrée moyenne dans le cas d'une contamination gaussienne.	49
2.4	Erreur absolue moyenne dans le cas d'une contamination de Student. . . .	60
2.5	Racine de l'erreur carrée moyenne dans le cas d'une contamination de Student.	61
2.6	Statistiques concernant le premier jeu de données.	62
2.7	Statistiques concernant le second jeu de données.	63
2.8	Comparaison de l'erreur de prédiction pour les deux jeux de données. . . .	64
2.9	MAE dans le cas d'une contamination de Student multivariée.	71
2.10	RMSE dans le cas d'une contamination de Student multivariée.	72
3.1	Carré moyen de l'erreur pour le cas où $\mu_0 = 0$ et $\sigma = 1$	105
3.2	Carré moyen de l'erreur pour le cas où $\mu_0 = 3$ et $\sigma = 4$	106
3.3	Estimations des paramètres du drap stable associés aux résidus du traitement géostatistique des données de Mercer and Hall (1911).	109
3.4	Intervalles de confiance pour les paramètres du drap.	110

Introduction

L'intérêt pour la statistique est dans la nature humaine. Dans la vie de tous les jours et quelque soit la situation face à laquelle un individu est confronté, il va chercher, consciemment ou non, à tirer de l'information des stimuli auxquels il est exposé. Que sa motivation soit de l'ordre de la curiosité, de la survie, ou de la cupidité, ce phénomène se justifie d'autant plus facilement que le nombre de stimuli augmente. Certains diront que ce ne sont que des balivernes puisqu'ils n'y connaissent rien en statistique; d'autres, que cette affirmation représente une vision tristement réductrice de cette science. Il serait aisé de démontrer aux premiers qu'ils se trompent. Quant aux seconds, j'ai l'espoir que le travail présenté ci-après saura leur prouver l'intérêt et le respect que je témoigne à cette branche des mathématiques dites appliquées.

La statistique, est une grande famille. L'ensemble de ses membres ont quelque chose en commun, mais il existe des groupes qui possèdent chacun leurs intérêts propres et qui apportent ainsi de la diversité. Cette dernière est fréquemment engendrée par les intérêts que l'on possède en dehors de la statistique. Tout du moins, je pense que c'est pour cela que je me suis orienté vers le groupe intitulé *Analyse de données spatiales*. Il existe comme pour tous les autres groupes une profusion d'ouvrages consacrés au sujet. On peut entre autre citer Cressie (1993) ou plus récemment Schabenberger and Gotway (2005) comme références globales sur le sujet. En bref, on peut dire que sous ce titre sont regroupées toutes les méthodes qui sont dédiées à l'étude et au traitement de données possédant une localisation spatiale. C'est-à-dire qu'en plus de la variable d'intérêt à proprement parler, on possède de l'information sur le lieu où cette variable a été observée. Bien évidemment, cela ne présente de l'intérêt que si le nombre de sites d'observations est multiple. Dans ce groupe, on peut assez facilement distinguer trois familles. Une première famille s'intitule *Configurations spatiales*. Ses membres s'intéressent principalement à la distribution dans l'espace du phénomène d'intérêt. Un exemple concret pourrait être les apparitions du loup sur le territoire national. Dans ce cas, les statisticiens pourraient chercher à savoir de

quelle manière ces apparitions sont distribuées sur le territoire. Si en plus des apparitions du loup, ils s'intéressent au nombre de spécimen observé ou à leur sexe, on parle dans ce cas de processus marqués. Une autre famille s'intitule *Données régionalisées*. Cette famille englobe tous ceux qui s'intéressent de plus près à des données observées non pas dans un espace continu, mais plutôt un espace composé d'une quantité dénombrable de régions bien définies et juxtaposées d'une certaine manière. Un exemple que tout citoyen a déjà eu sous les yeux : un exemplaire d'un quotidien national au lendemain d'une votation fédérale. Les statisticiens pourraient être intéressés, à partir des résultats cantonaux à extraire une tendance et quantifier une dépendance spatiale intercantonale. La dernière famille s'intitule *Données géostatistiques*. Ses membres considèrent que les données qu'ils ont en leur possession représentent uniquement la partie visible d'un processus. Ils supposent généralement que le processus lui-même existe de manière continue sur l'ensemble du domaine d'étude. Un exemple pourrait être l'élaboration du bulletin d'avalanche. A partir des observations récoltées par quelques dizaines de stations de mesures automatiques disséminées sur le territoire alpin national, il s'agit de prédire le danger d'avalanche sur l'ensemble du territoire. Finalement, la présentation du groupe *Analyse de données spatiales* ne serait pas complète si nos plus proches cousins n'étaient pas mentionnés. Leur groupe s'intitule *Analyse de séries temporelles*. La dépendance présente dans les données constitue notre point commun. Pour eux, cette dépendance est liée à un décalage temporel. Pour les géostatisticiens, cette dépendance est liée à un décalage mais également à une direction dans l'espace. Les relations sont d'autant plus forte que la tentation est grande de vouloir mélanger les deux domaines. On parle dans ce cas de modélisation spatio-temporelle. La présentation globale du domaine étant faite, le découpage de ce travail est proposé ci-dessous.

En géostatistique, les principaux thèmes sont la mesure de la dépendance spatiale et la prédiction. La dépendance spatiale est généralement quantifiée à l'aide d'une fonction de variogramme. De nombreuses variations existent tant dans les méthodes d'estimation que dans celles de l'ajustement de cette fonction. Concernant la prédiction, la méthodologie choisie est fortement dépendante de la modélisation choisie. Le Chapitre 1 présente un exposé relativement exhaustif des instruments de base de la géostatistique.

La décomposition d'un processus spatial en une somme d'un processus de tendance, d'un processus intrinsèquement stationnaire et d'un processus d'erreur est très souvent employée. La démarche est sensiblement toujours la même. A savoir l'estimation et la soustraction de la tendance aux données dans un premier temps ; puis l'ajustement du processus supposé intrinsèquement stationnaire. Lorsque l'on traite de tels processus, la qualité de la prédiction dépend principalement de la modélisation choisie. En effet, l'optimalité du krigeage ordinaire découle de la normalité supposée du processus sous étude. Malheureusement, il est fréquent qu'un processus physique possède une distribution dont

les queues sont plus lourdes que celles d'une distribution gaussienne. La nature linéaire du krigeage ordinaire fait qu'une faible proportion d'observations anormalement éloignées du centre de la distribution suffit à détruire cette optimalité (Huber, 1981). L'apparition de ce genre d'observations est une conséquence directe de la présence d'une distribution à queues lourdes. Pour ces modélisations dites à erreurs additives, des méthodes pour contrer l'absence de normalité du processus ont été proposées (Hawkins and Cressie, 1980). Il s'agit essentiellement d'un krigeage ordinaire appliqué à une version éditée des observations. La méthode peut également être vue comme une combinaison pondérée des données où la pondération traduit une mesure de fiabilité associée à l'observation elle-même.

Considérer des processus d'erreurs comme ceux énoncés précédemment revient à dire que ce que l'on a observé est une version perturbée du processus d'intérêt. Le Chapitre 2 propose une nouvelle modélisation pouvant servir à traiter des jeux de données présentant une hétérogénéité manifeste dans l'échelle. Il est cette fois-ci supposé que la contamination, lorsqu'elle apparaît, ne s'ajoute pas au processus d'intérêt, mais l'éclipse totalement. Autrement dit, une observation jugée contaminée ne véhicule aucune information pertinente. Par opposition à la modélisation présentée auparavant, celle-ci a été baptisée modélisation à erreurs substitutives. La loi de la contamination est choisie de sorte que ses queues soient plus lourdes que celles de la loi du processus d'intérêt. Le processus de localisation de la contamination est tout d'abord supposé dénué de dépendance spatiale. De plus, le processus de contamination est supposé, dans un premier temps, également gaussien. L'estimateur théorique optimal est dérivé et ses propriétés présentées. En se basant sur une méthode proposée par Cerioli and Riani (1999), une approximation de cet estimateur optimal est proposée et ses qualités sont éprouvées au travers de simulations. La méthode y est comparée à différentes méthodes de prédiction, classiques et robustes. Dans un second temps, la loi du processus de contamination est généralisée aux lois elliptiques. Des simulations sont également effectuées pour étayer la qualité de l'estimateur proposé et ainsi démontrer son indépendance vis-à-vis de la normalité de la distribution du processus de contamination. Une application à un jeu de données réelles et pseudo-réelles est proposée. Pour clore le chapitre, une variation de l'estimation est proposée pour pallier au relâchement de l'absence de dépendance spatiale dans le processus de localisation de la contamination.

Là où le krigeage robuste proposé par Hawkins and Cressie (1980) apparaît comme une version atténuée de la surface observée, la méthode de krigeage proposée dans le Chapitre 2, elle, retourne des surfaces semblables à celles que l'on pourrait observer en l'absence d'une telle contamination. Dès lors que l'hétérogénéité présente dans les données n'est plus en partie absorbée par l'ajustement, une conséquence est de retrouver cette hétérogénéité dans les résidus de prédiction. Le Chapitre 3 propose une nouvelle modélisation d'une surface pouvant exhiber cette caractéristique. Elle est intitulée drap stable. Le drap stable est

une généralisation d'un mouvement de Lévy unidimensionnelle où les incréments sont supposés suivre une distribution stable. Les distributions stables apparaissent naturellement lorsque l'on considère les distributions de sommes asymptotiques de variables aléatoires. Elles représentent en fait les seules lois limites envisageables ; la loi normale en fait partie et constitue simplement un cas très particulier, celui où le deuxième moment existe. Une particularité de ces lois est l'absence, en général, de version explicite de leur fonction de distribution. Une méthode d'ajustement des paramètres d'un tel processus stochastique est proposée. Il s'agit d'une approximation numérique par transformée de Fourier de la vraisemblance après relaxation des contraintes. Des simulations sont présentées pour juger de la qualité de l'estimateur proposé. Une application termine ce chapitre.

Une conclusion clos ce travail. La bibliographie qui la suit présente la littérature qui m'a permise de mettre sur pied ce travail.

CHAPITRE 1

Géostatistique

Le titre de ce chapitre, contraction du préfixe *géo* et du mot *statistique*, détermine tout une aire de la statistique moderne qui a pour but d'expliquer sinon d'étudier des processus aléatoires se réalisant sur un domaine spatial donné. La géostatistique est apparue pour la première fois par le biais de Matheron (1962) qui la définit et l'utilise comme un mélange d'ingénierie minière, de géologie, de mathématique et de statistique. Elle trouve ses racines dans l'exploitation minière ou pétrolière principalement où elle sert à étudier et cartographier le sol et ses ressources en fonction des prélèvements disponibles. Son champs d'application est vaste. En effet, elle s'applique couramment à l'étude de phénomènes de contamination, de pollution, de conductivité électrique du sol, de porosité ou de précipitations. En résumé, elle est principalement utilisée dans les études environnementales.

Penser que le champ d'application de la géostatistique est limité aux problèmes environnementaux serait cependant une erreur. A y réfléchir, n'importe quel jeu de données est pourvu d'un aspect spatial et/ou temporel. Exploiter cette dimension ou non est une question auquel tout statisticien se doit dans un premier temps de répondre. Les méthodes spatiales, comme leur nom l'indique, tiennent compte justement de cette dimension spatiale contrairement aux méthodes plus classiques. Cependant, même si bien souvent des données observées dans un voisinage (spatial ou temporel) proche partagent plus que des observations éloignées, le choix d'utiliser des méthodes incluant cette information relève du cas par cas et du choix de chacun. Finalement, le modèle classique, sans spécification spatiale peut, bien souvent, être vu comme un cas particulier du modèle spatial plus général.

D'ailleurs, si la structure moyenne ou tendance du processus dépend de différentes variables explicatives, l'emploi d'une méthode spatiale pour analyser la structure des résidus peut être considéré comme une protection contre un mauvais ajustement de la structure moyenne. Au vu de ces différents arguments, les méthodes géostatistiques apparaissent comme riches de potentiel et d'intérêt. Puisque le thème de ce travail est l'étude et le traitement de données spatiales, il est nécessaire, pour commencer, d'en présenter les bases. Les lignes qui suivent se veulent une présentation globale mais détaillée de la méthodologie spatiale classique.

1.1 Processus univarié

En toute généralité, un jeu de données spatiales $\{z(s_i) : s_i \in \mathcal{D}, i = 1, \dots, n\}$, où $s_i \in \mathcal{D} \subset \mathbb{R}^d$, $i = 1, \dots, n$ représentent les localisations des observations, peut être vu comme la réalisation d'un processus stochastique ou champs aléatoire noté

$$\{Z(s) : s \in \mathcal{D}\}. \quad (1.1)$$

L'ensemble \mathcal{D} est un volume d -dimensionnel généralement borné qui représente le domaine sur lequel se réalise le processus $Z(\cdot)$. Il peut être aléatoire ou pas. En géostatistique, il est généralement considéré comme fixe. D'autre part, la cardinalité de \mathcal{D} peut-être finie, dénombrable ou infinie. En géostatistique, on suppose généralement que la cardinalité de \mathcal{D} est infinie ; autrement dit, que l'indice s varie dans un espace continu et que ce que l'on a observé est en quelque sorte la partie visible d'un processus sous-jacent. Il faut encore préciser que si on voulait emphaser la nature aléatoire du processus, il conviendrait plutôt d'écrire le processus (1.1) comme $\{Z(s, \omega) : s \in \mathcal{D}; \omega \in \Omega\}$, où $(\Omega, \mathcal{F}, \mathbf{P})$ est un espace de probabilité. Ainsi, la réalisation $\{z(s_i) : s_i \in \mathcal{D}, i = 1, \dots, n\}$ correspondrait à une valeur particulière de ω .

Le processus (1.1) est défini par la loi conjointe

$$\mathbf{F}_{s_1, \dots, s_l}(z_1, \dots, z_l) = \mathbf{P}(Z(s_1) \leq z_1, \dots, Z(s_l) \leq z_l) \quad l \geq 1.$$

Cette loi est supposée vérifier les conditions de Kolmogorov, à savoir la condition de symétrie (la loi \mathbf{F} est invariante si l'on applique la même permutation à s_i et z_i) et celle de consistance ($\mathbf{F}_{s_1, \dots, s_k, s_{k+1}, \dots, s_l}(z_1, \dots, z_k, \infty, \dots, \infty) = \mathbf{F}_{s_1, \dots, s_k}(z_1, \dots, z_k)$). Si la loi conjointe est gaussienne, le processus défini par (1.1) est appelé processus gaussien. Pour que les données que l'on possède ne représentent pas uniquement un échantillon incomplet d'une unique réalisation, il est nécessaire de faire des suppositions supplémentaires sur le comportement du processus aléatoire en question. Les aspects en question sont présentés dans les deux sections suivantes.

1.1.1 Stationnarité

Définition 1.1 Un processus $Z(\cdot)$ est dit **strictement stationnaire** si

$$F_{s_1, \dots, s_l}(z_1, \dots, z_l) = F_{s_1+h, \dots, s_l+h}(z_1, \dots, z_l),$$

pour tout $l \in \mathbb{N}$, $s_1, \dots, s_l \in \mathcal{D}$, $z_1, \dots, z_l \in \mathbb{R}$ et tout $h \in \mathbb{R}^d$.

Définition 1.2 Un processus $Z(\cdot)$ est dit **stationnaire au $k^{\text{ème}}$ ordre** si les termes de l'égalité ci-dessous existent et vérifient

$$E(Z(s_1) \cdot \dots \cdot Z(s_l)) = E(Z(s_1+h) \cdot \dots \cdot Z(s_l+h)),$$

pour tout $1 \leq l \leq k$, $s_1, \dots, s_l \in \mathcal{D}$ et tout $h \in \mathbb{R}^d$.

La stationnarité au $k^{\text{ème}}$ ordre implique que les premiers moments, jusqu'au $k^{\text{ème}}$, existent et sont bornés. Clairement, la stationnarité stricte (ou forte) qui concerne la distribution du processus implique la stationnarité au $k^{\text{ème}}$ ordre pour tout $k \in \mathbb{N}$, pour autant que les moments existent. La réciproque n'est pas vraie.

Proposition 1.3 Un processus $Z(\cdot)$ est **stationnaire au second ordre** si les moments suivants existent et vérifient

$$\begin{aligned} E(Z(s_i)) &\equiv \mu, \quad \forall s_i \in \mathcal{D}, \\ \text{Cov}(Z(s_i), Z(s_j)) &= \mathcal{C}(s_i - s_j), \quad \forall s_i, s_j \in \mathcal{D}. \end{aligned}$$

La fonction $\mathcal{C}(\cdot)$ est appelée **covariogramme**. Si la fonction ne dépend pas de la direction, mais uniquement de la distance $\|s_1 - s_2\|$, elle est dite **isotropique**. Un processus $Z(\cdot)$ stationnaire au second ordre est également appelé processus **faiblement stationnaire**.

Comme le précise Cressie (1993), il est souvent suffisant, pour effectuer une prédiction spatiale par exemple, d'être en présence d'une stationnarité légèrement plus faible que celle du second ordre.

Définition 1.4 Un processus $Z(\cdot)$ qui satisfait

$$\begin{aligned} E(Z(s_i)) &\equiv \mu, \quad \forall s_i \in \mathcal{D}, \\ \text{Var}(Z(s_i) - Z(s_j)) &= 2\gamma(h = s_i - s_j), \quad \forall s_i, s_j \in \mathcal{D}. \end{aligned}$$

est dit **intrinsèquement stationnaire**. La fonction $2\gamma(\cdot)$ est appelée **variogramme**. Si cette fonction ne dépend pas de la direction, mais uniquement de la distance $h = \|s_1 - s_2\|$,

elle est dite isotropique.

Tout processus qui est stationnaire au second ordre est également intrinsèquement stationnaire. La réciproque n'est pas vraie. Pour qu'elle le soit, il faut encore imposer que le variogramme $2\gamma(h)$ soit borné pour tout $h \in \mathbb{R}^d$. Le variogramme tient une place importante en géostatistique. Il en sera davantage question à la Section 1.1.3. La plupart des processus étudiés sont supposés présenter une autre importante propriété, à savoir l'**ergodicité**. C'est ce de quoi il est question à la section suivante.

1.1.2 Ergodicité

Malgré les apparences, le fait que l'on aie observé l'échantillon $\{z(s_i) : s_i \in \mathcal{D}, i = 1, \dots, n\}$ ne signifie pas que la taille de l'échantillon soit n . En effet, un processus stochastique étant en fait une famille de variables aléatoires, la taille de l'échantillon observé est en fait égale à un et correspond, à un événement particulier $\omega = \omega_0 \in \Omega$. À partir de là, si l'on souhaite pouvoir tout de même effectuer de l'inférence, basée sur un unique ensemble d'observations dépendantes, on a besoin de la notion d'ergodicité. Pour une définition de l'ergodicité, voir Adler (1981). De manière succincte, cette propriété permet d'estimer l'espérance sur l'ensemble fondamental Ω par la moyenne empirique spatiale $\bar{Z} = \sum_i Z(s_i)/n$ ou la dépendance spatiale par l'estimateur fourni par (1.9).

Pourtant, pour se convaincre de la puissance de la propriété d'ergodicité d'un processus, la remarque suivante devrait faire l'affaire. Fréquemment, seule la stationnarité au second ordre est supposée pour des processus spatiaux. Le problème est que si un processus stationnaire au second ordre est supposé ergodique, il doit forcément être strictement stationnaire. Donc, contrairement à ce que l'on pourrait penser, il n'y a dans ce cas de généralité à gagner en supposant uniquement une stationnarité au second ordre (Cressie, 1989). En fait, les statisticiens utilisent rarement la propriété d'ergodicité sous d'autres aspects que ceux qui leur assurent une consistance (convergence en L_2) de la moyenne empirique et de l'estimateur de la fonction de covariance. Cette forme d'ergodicité est dite faible et dans ce cas, il est évident que la stationnarité au second ordre est une condition plus faible que la stationnarité stricte. C'est cette ergodicité qui sera supposée tout au long de ce travail.

Pour des processus gaussiens stationnaires au second ordre, la condition que

$$\lim_{h \rightarrow \infty} \mathcal{C}(h) = 0$$

suffit pour avoir un processus ergodique. Une idée serait donc, à partir d'un jeu de données supposées gaussiennes, d'estimer $\mathcal{C}(h)$ par $\hat{\mathcal{C}}(h)$ et de vérifier visuellement, à l'aide d'un graphique, si l'estimation obtenue vérifie bien cette condition. Le problème est que l'ergodicité est déjà nécessaire pour que $\hat{\mathcal{C}}(h)$ soit un estimateur consistant. Il est donc impossible

d'attester à partir de données si un processus est oui ou non ergodique.

1.1.3 Variogramme

Définition 1.5 Supposons que

$$\text{Var}(Z(s_1) - Z(s_2)) = 2\gamma(s_1 - s_2), \quad \forall s_1, s_2 \in \mathcal{D}.$$

La fonction $2\gamma(\cdot)$ est appelée **variogramme**. Cette fonction est définie négative ; c'est-à-dire qu'elle vérifie

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j 2\gamma(s_i - s_j) \leq 0, \quad (1.2)$$

pour tout choix fini de localisations $\{s_i : i = 1, \dots, m\}$ et de nombres réels $\{a_i : i = 1, \dots, m\}$ satisfaisant la condition $\sum_{i=1}^m a_i = 0$. Elle vérifie les propriétés

$$\gamma(h) = \gamma(-h), \quad \gamma(0) = 0, \quad \lim_{\|h\| \rightarrow \infty} \frac{\gamma(h)}{\|h\|^2} = 0.$$

Définition 1.6 Si le variogramme est tel que

$$\lim_{h \rightarrow 0} \gamma(h) = \theta_1 > 0,$$

θ_1 est appelé l'**effet pépité**.

La présence d'un effet pépité est en contradiction avec l'hypothèse que le processus $Z(\cdot)$ soit L_2 -continu. Par conséquent, si on suppose la continuité, on peut conclure que θ_1 représente l'erreur de mesure. Comme en pratique, on ne connaît rien de la variation du processus à une échelle inférieure à $\min\{\|s_i - s_j\| : 1 \leq i, j \leq n\}$ on ne peut rien affirmer sur la vraie nature du processus. Il est donc plus avisé de modéliser le processus comme la somme d'un processus L_2 -continu et d'un bruit blanc. L'effet pépité peut ainsi être vu comme la somme de l'erreur de mesure et de la variance du bruit blanc. L'effet pépité est généralement estimé par extrapolation à partir des valeurs prises par le variogramme au voisinage de 0. On peut énoncer les propriétés suivantes à propos du comportement du variogramme autour de l'origine.

Propriété 1.7

- Si $2\gamma(\cdot)$ est continu à l'origine, alors $Z(\cdot)$ est L_2 -continu.
- Si $2\gamma(\cdot)$ est discontinu à l'origine, alors $Z(\cdot)$ est également discontinu et hautement irrégulier.

En général, on suppose que le processus possède un variogramme paramétrisé. A partir des données, il faut donc déterminer la forme générale ainsi que la valeur des différents

paramètres. Les paramètres sont au nombre de trois :

θ_1 : l'effet pépité ;

θ_2 : le **seuil** qui correspond à la valeur maximale prise par le variogramme auquel on soustrait l'effet pépité ;

θ_3 : un paramètre proportionnel à l'**étendue** du variogramme. L'étendue est définie comme étant la valeur pour laquelle le variogramme atteint son maximum dans une direction donnée. En cas de processus isotropique, cette valeur est unique.

Ci-dessous, quelques variogrammes isotropiques parmi les plus fréquemment utilisés, paramétrisés en fonction de $\theta = (\theta_1, \theta_2, \theta_3)^t$. Ces différentes fonctions sont représentées dans la Figure 1.1.

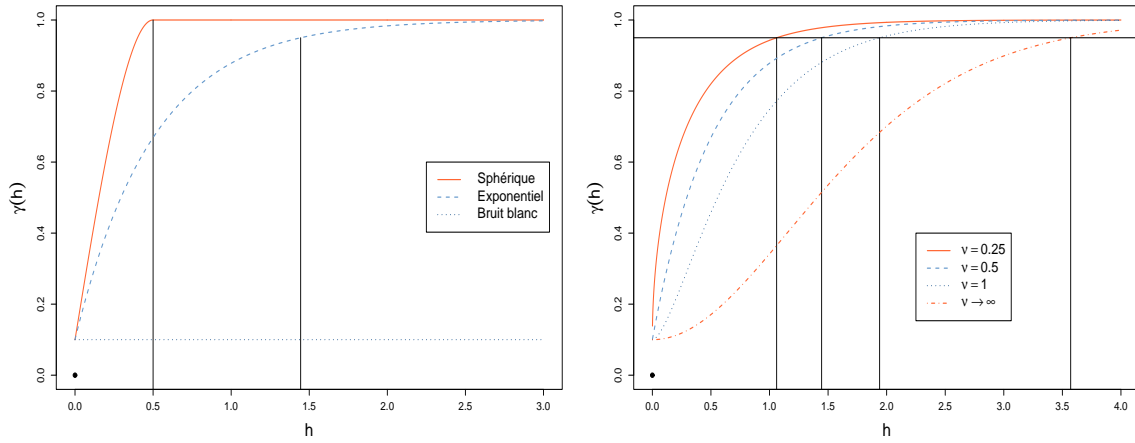


Figure 1.1. Représentations de variogrammes isotropiques avec $\theta = (0.1, 0.9, 0.5)^t$. Dans le graphique de gauche, le variogramme, respectivement sphérique (1.4), exponentiel (1.5) et bruit blanc (1.3) est représenté. On remarque que l'étendue du variogramme sphérique est finie et égale à θ_3 ; l'étendue pratique du variogramme exponentiel est également mise en évidence. A droite, des variogrammes de Matérn (1.6), pour différentes valeurs du paramètre de lissage ν . Le cas $\nu = 0.5$ correspond au variogramme exponentiel dans le graphique de gauche. Les différentes étendues pratiques sont mises en évidence. Clairement, l'étendue pratique augmente avec ν .

Bruit blanc :

$$\gamma(h; \theta_1, \theta_2, \theta_3) = \begin{cases} 0, & \text{si } \|h\| = 0, \\ \theta_1, & \text{sinon.} \end{cases} \quad (1.3)$$

Ce modèle correspond à un processus de bruit blanc. Autrement dit, pour ce processus, $\text{Cov}(Z(s_i), Z(s_j)) = 0$, $\forall s_i, s_j \in \mathcal{D}$, $s_i \neq s_j$.

Variogramme sphérique :

$$\gamma(h; \theta_1, \theta_2, \theta_3) = \begin{cases} 0, & \text{si } \|h\| = 0, \\ \theta_1 + \theta_2 \left(\frac{3}{2} \frac{\|h\|}{\theta_3} - \frac{1}{2} \left(\frac{\|h\|}{\theta_3} \right)^3 \right), & \text{si } 0 < \|h\| \leq \theta_3, \\ \theta_1 + \theta_2, & \text{sinon.} \end{cases} \quad (1.4)$$

Un des atouts qui rend ce variogramme intéressant est qu'il possède une étendue finie. Autrement dit, passé une certaine distance $h > \theta_3$, il n'y a plus de dépendance spatiale.

Variogramme exponentiel :

$$\gamma(h; \theta_1, \theta_2, \theta_3) = \begin{cases} 0, & \text{si } \|h\| = 0, \\ \theta_1 + \theta_2 (1 - \exp(-\|h\|/\theta_3)) & \text{sinon.} \end{cases} \quad (1.5)$$

De part sa nature exponentielle, il possède une étendue infinie.

Variogramme de Matern :

$$\gamma(h; \theta_1, \theta_2, \theta_3) = \begin{cases} 0, & \text{si } \|h\| = 0, \\ \theta_1 + \theta_2 \left(1 - 2\Gamma(\nu)^{-1} \left(\frac{\|h\|}{2\theta_3} \right)^\nu K_\nu(\|h\|/\theta_3) \right), & \text{sinon,} \end{cases} \quad (1.6)$$

où $K_\nu(\cdot)$ représente la fonction de Bessel de deuxième espèce d'ordre $\nu > 0$. Le paramètre ν peut être vu comme un paramètre de lissage. En particulier, lorsque $\nu = 1/2$, on obtient le variogramme exponentiel et lorsque $\nu \rightarrow \infty$, on obtient le variogramme dit gaussien (Schabenberger and Gotway, 2005).

Etant donné que pour certaines fonctions de variogramme l'étendue est infinie, on parle souvent d'étendue pratique. Elle correspond à la valeur pour laquelle le variogramme atteint les 95% de sa valeur maximale. La relation entre le variogramme $2\gamma(\cdot)$ et le covariogramme $\mathcal{C}(\cdot)$ défini dans la Proposition 1.3 est donnée par l'équation

$$\text{Var}(Z(s_1) - Z(s_2)) = \text{Var}(Z(s_1)) + \text{Var}(Z(s_2)) - 2\text{Cov}(Z(s_1), Z(s_2)). \quad (1.7)$$

Autrement dit, si on considère un processus stationnaire au second ordre, la relation (1.7) nous fournit directement un variogramme puisque l'on peut réécrire

$$2\gamma(s_1 - s_2) = 2(\mathcal{C}(0) - \mathcal{C}(s_1 - s_2)). \quad (1.8)$$

Comme écrit auparavant, on en déduit que tout processus stationnaire au second ordre est également intrinsèquement stationnaire. On peut même dire que la classe des pro-

cessus stationnaire au second ordre est strictement contenue dans celle des processus intrinsèquement stationnaires puisqu'il existe des processus intrinsèquement stationnaires qui ne sont pas stationnaires au second ordre, voir par exemple Cressie (1993, p.68).

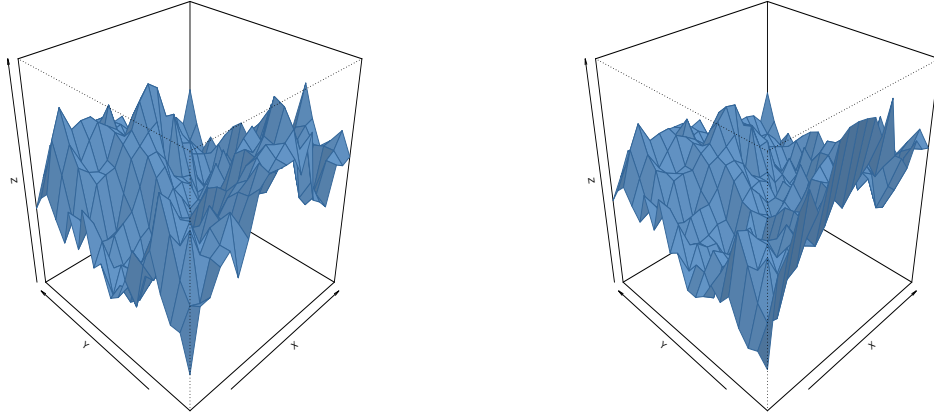


Figure 1.2. Représentations de deux réalisations de processus intrinsèquement stationnaires sur une grille régulière couvrant le carré unitaire $\mathcal{D} = [0, 1] \times [0, 1]$ obtenues par simulation. Ici, $E(Z(s_i)) = \mu = 0$, $\forall s_i \in \mathcal{D}$ pour les deux processus. Dans les deux cas, le variogramme considéré est sphérique et isotropique. À gauche, l'étendue vaut $\theta_3 = 0.5$ et à droite l'étendue vaut $\theta_3 = 1.5$. On constate que la dépendance spatiale possédant une plus forte étendue se traduit par l'obtention d'une surface plus lisse (à droite).

1.1.4 Anisotropie

Si la structure spatiale est isotropique, les courbes de niveaux de la fonction de covariance $\mathcal{C}(h)$ sont sphériques. Par contre, si la fonction de covariance est anisotropique, cela signifie que la structure spatiale est dépendante de la direction considérée et ses courbes de niveaux peuvent prendre des formes beaucoup plus générales. Il y a toutefois un cas particulier ; le cas où les courbes de niveaux adoptent des contours elliptiques. D'après Matérn (1986), ce cas est appelé anisotropie géométrique et peut être corrigé à l'aide d'une transformation linéaire du système de coordonnées. En pratique, il apparaît fréquemment lorsque l'on traite des processus qui évoluent selon des directions préférentielles, comme par exemple une pollution de l'air qui subit généralement l'effet directionnel du vent. Comme décrit par Schabenberger and Gotway (2005), la méthode consiste à :

- I :** Appliquer une rotation au système de coordonnées pour qu'il s'aligne avec le petit axe et le grand axe de l'ellipse.
- II :** Effectuer une homothétie dans la direction du grand axe afin de le compresser et ainsi rendre les contours sphériques.

Cette démarche revient à appliquer la matrice suivante au système de coordonnées,

$$A = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

où λ détermine le rapport d'anisotropie. Pratiquement, le variogramme présente la même forme générale et le même seuil dans les deux directions correspondant aux deux axes principaux, mais différent dans leurs valeurs respectives de l'étendue. Le rapport d'anisotropie est égal au rapport des deux étendues respectives. Dans la section suivante sont présentées les méthodes classiques d'estimation et d'ajustement du variogramme.

1.2 Estimation et ajustement de la variation spatiale

Comme nous l'avons vu précédemment, la structure du second moment de notre processus $Z(\cdot)$ dépend de ses propriétés et en particulier du type de stationnarité qu'il exhibe. Si le processus étudié est stationnaire au second ordre, un covariogramme peut être estimé et ajusté. Si il est uniquement intrinsèquement stationnaire, un variogramme peut être estimé. Il existe une multitude d'estimateurs ; nous nous bornerons à présenter ici les plus classiques d'entre eux. C'est ceux que l'on retrouve dans des ouvrages de références comme Cressie (1993) ou, plus récemment, Schabenberger and Gotway (2005). Certains d'entre eux possèdent de bonnes propriétés théoriques, d'autres moins. C'est pourquoi, après avoir obtenu une estimation de la variation spatiale, on ajuste généralement un modèle théorique dont on connaît les propriétés à l'estimation obtenue. Les deux prochaines sections détaillent ces propos.

1.2.1 Estimation

Une idée relativement intuitive pour étudier la variabilité spatiale est de représenter la différence carrée $(Z(s_i) - Z(s_j))^2$ en fonction du décalage h . Le problème est que pour un décalage fixé, le nombre de paires à considérer peut être passablement élevé et le graphique est en général pas vraiment informatif. D'où l'idée de résumer pour chaque valeur du décalage l'ensemble des carrés des différences par leur moyenne. Cet estimateur, proposé par Matheron (1962) est appelé l'estimateur classique du variogramme et s'écrit

comme

$$2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} (Z(s_i) - Z(s_j))^2, \quad (1.9)$$

où

$$N(h) = \{(s_i, s_j) : s_i - s_j \in V(h)\} \quad (1.10)$$

et $V(h)$ représente un voisinage dans \mathbb{R}^d . On considère un voisinage ou une région de tolérance, car si les sites ne sont pas régulièrement espacés (la majorité des cas), le nombre de couples tels que leur différence égale exactement h est faible. Ce voisinage peut être par exemple défini comme $V(h; a, b) = \{d : \|d - h\| \leq a \text{ et } \arccos(d \cdot h / \|d\| \cdot \|h\|) \leq b\}$. En plus de considérer un voisinage ou une région de tolérance des règles pratiques sont admises. Par exemple, le degré de tolérance devrait être suffisant pour qu'au moins trente paires soient disponibles dans chaque voisinage ou le variogramme est estimé; de même, le variogramme ne devrait pas être estimé à un décalage h tel que $\|h\| > D/2$ où D est le diamètre de \mathcal{D} . Les propriétés qui font de cet estimateur le plus utilisé en pratique sont son absence de biais, sa parité et le fait qu'il prenne la valeur 0 à un décalage $h = 0$. Malheureusement, de part sa nature (somme de carrés de différences), l'estimateur classique (1.9) qui correspond à l'estimateur des moments ne présente aucune résistance aux valeurs aberrantes. Non seulement la déviation entre le site contaminé et les autres sites sera faussement amplifiée, mais, de plus, puisque un site entre dans l'estimation du variogramme pour différents décalages, la contamination sera diffusée sur l'ensemble de l'estimation. Dès lors, la robustesse aux valeurs aberrantes est apparue comme une propriété cruciale lors du développement d'autres estimateurs.

Cressie and Hawkins (1980b) ont développé un estimateur qu'ils ont qualifié de robuste. La base de son élaboration repose sur la constatation que la racine quatrième de la différence carrée $(Z(s_i) - Z(s_j))^2$ se comporte approximativement de manière gaussienne. Ils proposent l'estimateur

$$2\hat{\gamma}(h) = \frac{1}{b(N(h))} \left(\frac{1}{|N(h)|} \left(\sum_{N(h)} |Z(s_i) - Z(s_j)|^{1/2} \right) \right)^4, \quad (1.11)$$

où $N(h)$ est défini comme dans (1.10) et $b(N(h)) = 0.457 + 0.494/|N(h)|$ est un terme correctif pour le biais. Il faut préciser que cet estimateur a été qualifié de robuste de ses auteurs, car il exhibe de meilleures propriétés que l'estimateur classique (1.9) lorsque l'on considère un processus comportant une erreur additive. Si on considère une autre mesure de robustesse tel que le point de rupture (Hampel, 1971), cet estimateur ne mérite pas son qualificatif puisque il partage le même point de rupture nul que l'estimateur classique. C'est justement cette mesure de robustesse qui a motivé Genton (1998a) à utiliser l'estimateur d'échelle Q_n développé par Rousseeuw and Croux (1993) pour construire

l'estimateur

$$2\hat{\gamma}(h) = \left(2.2191 \{ |(Z(s_i + h) - Z(s_i)) - (Z(s_j + h) - Z(s_j))| : i < j \}_{(\lfloor N(h)/2 \rfloor + 1)} \right)^2, \quad (1.12)$$

où $N(h)$ est donné par (1.10). Puisque l'estimateur d'échelle Q_n possède un point de rupture maximal de 50%, l'estimateur (1.12) possède le même point de rupture concernant le processus des différences $Z(\cdot + h) - Z(\cdot)$, mais pas forcément concernant le processus $Z(\cdot)$. Genton (2001) a montré à l'aide de simulations que le point de rupture de (1.12) en terme du processus $Z(\cdot)$ s'élevait à environ 30%. Finalement, il est bon de préciser que les différents estimateurs présentés ci-dessus sont biaisés si le processus $Z(\cdot)$ n'est pas intrinsèquement stationnaire.

Concernant l'estimation du covariogramme, on peut utiliser l'estimateur des moments

$$\hat{C}(h) = \frac{1}{|N(h)|} \sum_{N(h)} (Z(s_i) - \bar{Z})(Z(s_j) - \bar{Z}),$$

où $N(h)$ est donné par (1.10) et $\bar{Z} = \sum_i Z(s_i)/n$. Si l'on souhaite éviter le biais de cet estimateur, on peut tirer avantage de l'équation (1.8) et estimer le covariogramme à partir d'un estimateur du variogramme comme

$$\hat{C}(h) = \left(\lim_{\|h\| \rightarrow \infty} \hat{\gamma}(h) \right) - \hat{\gamma}(h).$$

Une fois le variogramme estimé, il reste que sa valeur n'est connue que pour un nombre finis de décalages. Une possibilité d'obtenir une valeur du variogramme pour n'importe quel décalage serait, par exemple, de lisser notre estimation à l'aide d'une fonction noyau. Le problème est que le variogramme que l'on obtiendrait par ce biais ne satisferait probablement pas la condition de définie négativité, donnée par (1.2). Dans la Section suivante, les méthodes d'ajustement les plus classiques sont présentées.

1.2.2 Ajustement

Les différentes approches existantes peuvent être divisées en deux classes, celles dites paramétriques et celles dites non paramétriques. En général, ce sont plutôt celles entrant dans la première catégorie qui sont employées et ce sont ces mêmes approches qui sont ici brièvement présentées. Cette catégorie peut à son tour être subdivisées en deux familles de méthodes; les méthodes complètement paramétriques et les méthodes semi-paramétriques.

Les méthodes paramétriques sont celles qui utilisent directement les données (par opposition à celles qui utilisent des pseudo données, autrement dit des transformation

des données) et supposent qu'il existe une distribution sous-jacente. Les méthodes du maximum de vraisemblance ou du maximum de vraisemblance restreint appartiennent à ce groupe. Ces méthodes sont bien connues, mais leurs principaux défauts également. L'optimisation ne se fait pas toujours sans problème et les estimateurs de la structure de second ordre peuvent être sérieusement biaisés. Pour plus de détails, voir par exemple Cressie (1993, Section 2.6.3).

Les autres méthodes, dites semi-paramétrique, sont les plus fréquemment utilisées. On suppose que le variogramme estimé $\hat{\gamma}(h)$ provient d'un modèle paramétrique $\gamma(h; \theta)$ où θ contient l'ensemble des paramètres du modèle et on ajuste le variogramme en optimisant un critère d'optimalité. Le plus fréquemment utilisé est certainement le critère de la somme des moindres carrés ordinaires ; le vecteur des paramètres θ est estimé comme

$$\hat{\theta} = \arg \min_{\theta} \sum_{k=1}^K (\hat{\gamma}(h_k) - \gamma(h_k; \theta))^2,$$

où h_k , $k = 1, \dots, K$, représentent les différents décalages pour lesquels on possède une estimation du variogramme. Le problème de cette méthode est qu'elle ne tient pas compte des dépendances existantes entre les estimations du variogramme obtenues à différents décalages. Une manière d'y remédier est de choisir θ comme

$$\hat{\theta} = \arg \min_{\theta} \sum_{k,l=1}^K (\hat{\gamma}(h_k) - \gamma(h_k; \theta)) V_{k,l} (\hat{\gamma}(h_l) - \gamma(h_l; \theta)),$$

où la matrice V est l'inverse de

$$R(\theta) = \text{Cov}(\hat{\gamma}(h_k) - \gamma(h_k; \theta), \hat{\gamma}(h_l) - \gamma(h_l; \theta))_{1 \leq k, l \leq K}.$$

Cette méthode, dite de la somme des moindres carrés généralisée, présente une faiblesse. Celle de devoir obtenir une bonne estimation de $R(\theta)$. Cressie (1985) a proposé d'approcher les termes diagonaux de cette matrice de covariance par

$$\text{Var}(\hat{\gamma}(h_k)) \approx 2 \frac{\gamma(h_k; \theta)^2}{|N(h_k)|}, \quad k = 1, \dots, K \quad (1.13)$$

et d'approcher (faiblement) le problème de la somme des moindres carrés généralisés par celui de la somme des moindres carrés pondérés en remplaçant $R(\theta)$ par $W(\theta)$, une matrice diagonale dont les termes diagonaux sont donnés par (1.13). Cela revient à choisir θ comme

$$\hat{\theta} = \arg \min_{\theta} \sum_{k=1}^K N(h_k) \left(\frac{\hat{\gamma}(h_k)}{\gamma(h_k; \theta)} - 1 \right)^2. \quad (1.14)$$

Zimmerman and Zimmerman (1991) ont montré à l'aide de simulations Monte-Carlo que les méthodes d'estimation de la somme des moindres carrés ordinaires ou pondérés atteignent des performances similaires. En fait, il y a beaucoup plus à perdre à considérer les termes hors-diagonaux nuls. Genton (1998b) a développé une méthode dite de la somme des moindres carrés généralisés explicite. En supposant la normalité des observations ou du moins une distribution elliptique (Genton, 2000) et en partant du fait que l'estimateur classique (1.9) peut être vu comme une forme quadratique, une forme explicite de $R(\theta)$ est dérivée. La valeur optimale de θ est dérivée à l'aide d'une méthode non linéaire itérative. Pour conclure, Zimmerman and Zimmerman (1991) ont également montré que les méthodes les plus simples ne sont pas complètement dépassées en comparaison avec les méthodes plus compliquées et plus coûteuses en temps de calcul.

Le but d'ajuster la dépendance spatiale d'un processus aléatoire réside principalement dans le fait qu'elle représente l'élément principal de toute prédiction spatiale de ce même processus. La section suivante présente ce sujet.

1.3 Krigeage

Parmi les nombreuses méthodes de prédiction spatiale existantes, il en existent un certain nombre qui partagent en tout les cas partiellement leur dénomination. Ce sont les méthodes de **krigeage** (ou “*kriging*” en anglais). La méthode du krigeage possède une genèse trouble. Cela est sûrement dû au fait que ses motivations sont naturelles et communes à plusieurs domaines de la statistique. En effet, qu'on aborde le problème sous l'oeil de la géostatistique, des séries temporelles, du multivarié ou de la prospection minière, le problème est toujours de prédire de manière optimale le processus d'intérêt là où on ne l'a pas observé (ou à un instant où on ne l'a pas observé) à partir de données dépendantes. Cressie (1990) propose de mettre de l'ordre dans l'émergence de ces techniques (en particulier, en géostatistique). Tout d'abord, sur le plan terminologique, tout le monde s'entend pour dire que le mot krigeage est un hommage à Daniel G. Krige, un ingénieur sud-africain qui développa des méthodes empiriques pour déterminer la distributions de filons orifères dans les années cinquante. Sous des considérations plus méthodologiques, c'est Matheron (1962) qui l'a présenté comme tel.

Il s'agit de dériver le prévisseur linéaire sans biais optimal au sens du carré moyen de l'erreur pour $Z(s_0)$, la valeur que prend le processus $Z(\cdot)$ en une localisation s_0 . Autrement dit, on cherche les poids optimaux $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)^t$ tels que

$$E \left(\left(Z(s_0) - \sum_{i=1}^n \lambda_i Z(s_i) \right)^2 \right)$$

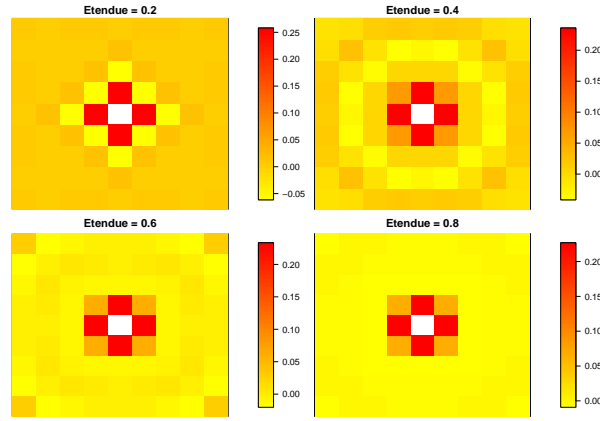


Figure 1.3. Quatre représentations de la valeur des poids optimaux du krigeage ordinaire donnés par (1.15) sur une grille régulière de $n = 49$ points couvrant le carré unitaire. La prédiction a lieu au centre de la grille. Le variogramme considéré est exponentiel avec un seuil égal à 1 et un effet pépité égal à 0.05. L'étendue est le seul paramètre qui varie dans les quatre cas de figure.

soit minimale. Pour que l'estimateur dérivé soit sans biais, il faut encore imposer que la somme des poids soit unitaire. On se retrouve par conséquent avec un problème de minimisation sous contrainte qui se résoud à l'aide de la méthode des multiplicateurs de Lagranges. Il faut préciser que ce problème peut également être résolu pour un processus stationnaire à l'ordre un seulement si son second moment est fini. Il suffit de définir le variogramme comme $2\gamma(s_i, s_j) = \text{Var}(Z(s_i) - Z(s_j))$ et le covariogramme comme $\mathcal{C}(s_i, s_j) = \text{Cov}(Z(s_i), Z(s_j))$ pour tout couple de localisations $s_i, s_j \in \mathcal{D}$ pour que les poids optimaux s'écrivent

$$\lambda = \Gamma^{-1} \left(\gamma + \mathbf{1} \frac{1 - \mathbf{1}^t \Gamma^{-1} \gamma}{\mathbf{1}^t \Gamma^{-1} \mathbf{1}} \right), \quad (1.15)$$

où

$$\begin{cases} \gamma = (\gamma(s_0 - s_1), \dots, \gamma(s_0 - s_n))^t, \\ \mathbf{1} = (1, \dots, 1)^t \in \mathbb{R}^n, \\ \Gamma = (\gamma(s_i - s_j))_{1 \leq i, j \leq n}. \end{cases}$$

Une représentation spatiale de ces poids optimaux est donnée par la Figure 1.3. Le prévisseur linéaire optimal sans biais ou **krigeage ordinaire** s'écrit

$$\hat{Z}(s_0) = \gamma^t \Gamma^{-1} Z + \frac{1 - \gamma^t \Gamma^{-1} \mathbf{1}}{\mathbf{1}^t \Gamma^{-1} \mathbf{1}} (\mathbf{1}^t \Gamma^{-1} Z), \quad (1.16)$$

où $Z = (Z(s_1), \dots, Z(s_n))$ et la variance de krigeage, ou carré moyen de l'erreur de prédiction, vaut

$$\sigma^2(s_0) = \gamma^t \Gamma^{-1} \gamma - \frac{(\mathbf{1}^t \Gamma^{-1} \gamma - 1)^2}{\mathbf{1}^t \Gamma^{-1} \mathbf{1}}.$$

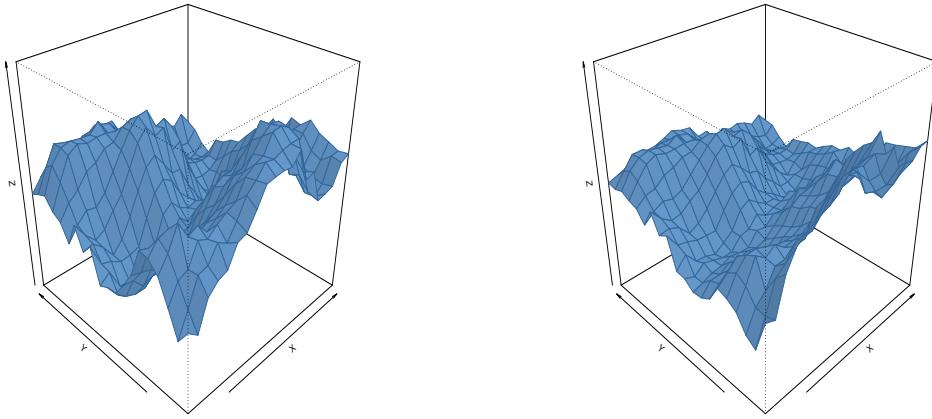


Figure 1.4. Représentations des surfaces obtenues après avoir appliqué le krigeage ordinaire aux deux surfaces de la Figure 1.2. Dans chacun des cas, la dépendance spatiale, autrement dit, la matrice Γ n'a pas été estimée mais fournie. La prédiction du processus possédant une plus forte dépendance spatiale (à droite) apparaît également plus lisse que l'autre. Puisque les sites de prédiction sont les mêmes que les sites d'observation, lors de chaque prédiction, l'observation en question a été omise. Sans quoi, on obtiendrait une interpolation exacte. Une autre manière de faire aurait été de considérer d'autres sites de prédiction comme, par exemple, la même grille mais légèrement décalée.

La Figure 1.4 présente des surfaces obtenues après application du krigeage ordinaire aux données. Si l'espérance $\mu(\cdot)$ et la structure de covariance $\mathcal{C}(\cdot, \cdot)$ sont connues, le prévisseur donné par (1.16) se simplifie et s'écrit

$$\hat{Z}(s_0) = \mu(s_0) + c^t C^{-1}(Z - \mu), \quad (1.17)$$

où $c = (\mathcal{C}(s_0, s_1), \dots, \mathcal{C}(s_0, s_n))^t$, $C = (\mathcal{C}(s_i, s_j))$ et $\mu = (\mu(s_1), \dots, \mu(s_n))^t$. Le prévisseur donné par (1.17) est dit **krigeage simple**. Si le processus $Z(\cdot)$ est gaussien, il coïncide avec l'espérance conditionnelle $E(Z(s_0) | Z(s_1), \dots, Z(s_n))$. Par conséquent, il est optimal au sens du carré moyen de l'erreur parmi tous les estimateurs sans biais, linéaires ou non.

Dans le cas où la structure moyenne n'est pas constante mais peut être écrite comme

$$\mu(s_i) = \sum_{j=0}^p \beta_j f_j(s_i), \quad \forall s_i \in \mathcal{D}, \quad (1.18)$$

où $\{f_j(\cdot) : j = 0, \dots, p\}$ sont des fonctions connues, le krigeage ordinaire peut être généralisé au **krigeage universel**. Essentiellement, cela revient à remplacer dans l'expression du krigeage ordinaire l'écriture $\mu \mathbf{1}$ par $X\beta$. La $j^{\text{ème}}$ colonne de X est donnée par les valeurs prises par la fonction $f_j(\cdot)$ dans les localisations s_1, \dots, s_n . D'autres méthodes

de krigeage ont été développées pour traiter différentes modélisations du processus spatial étudié. Par exemple, le **krigeage robuste** de Hawkins and Cressie (1980) présenté ci-dessous ou le krigeage par *median polish* sont deux méthodes proposées pour traiter des processus non stationnaires.

1.3.1 Krigeage robuste

L'optimalité du krigeage ordinaire en tant que prédiction spatiale linéaire repose entièrement sur la stationnarité au second ordre du processus. Par contre, son optimalité parmi l'ensemble des prévisseurs, linéaires ou non, dépend elle de la nature gaussienne du processus. Même si il existe des méthodes pour transformer les observations afin qu'elles satisfassent autant que possible à cette condition, il reste toujours des situations où apparaissent des observations typiquement mieux modélisées en tant que valeurs aberrantes. La première solution qui consiste à ne pas considérer ces observations à la nature douteuse peut encore présenter du sens lorsqu'il s'agit d'obtenir une estimation du variogramme (encore que différentes alternatives on été présentées dans la Section 1.2), mais quand il s'agit d'effectuer une prédiction, un moyen de travailler en leur présence s'avère nécessaire. De plus, les valeurs qui se distinguent par leur amplitude anormale ne sont pas tout le temps dénuées d'information sur le processus étudié. Par exemple, en prospection minière, un pic traduisant la présence d'un filon peut être synonyme de richesse et il serait malvenu de simplement le supprimer. Hawkins and Cressie (1980) ont proposé une méthode qui permet simplement d'atténuer l'influence de telles valeurs dans leur voisinage lorsque une prédiction est effectuée. La méthode proposée se décompose en cinq étapes.

I : Estimer de manière robuste la dépendance spatiale et ajuster un variogramme valide.

II : Appliquer le krigeage ordinaire aux observations afin d'obtenir pour chaque $Z(s_i)$ les poids λ_{il} , $l = 1, \dots, n$, $l \neq i$ tels que

$$\hat{Z}_{-i}(s_i) = \sum_{\substack{l=1 \\ l \neq i}}^n \lambda_{il} Z(s_l)$$

et la variance de krigeage associée $\sigma_{-i}^2(s_i)$.

III : Utiliser les poids obtenus à l'étape précédente pour obtenir une prédiction robuste de $Z(s_i)$:

$$Z_{-i}^{mp}(s_i) = \text{médiane pondérée} \left(\{Z(s_l) : l \neq i\}, \{\lambda_{li} : l \neq i\} \right). \quad (1.19)$$

IV : Editer l'observation en la substituant par une version *winsorisée* :

$$Z^e(s_i) = \begin{cases} Z_{-i}^{mp}(s_i) + c\sigma_{-i}(s_i) & \text{si } Z(s_i) - Z_{-i}^{mp}(s_i) > c\sigma_{-i}(s_i), \\ Z_{-i}^{mp}(s_i) - c\sigma_{-i}(s_i) & \text{si } Z(s_i) - Z_{-i}^{mp}(s_i) < -c\sigma_{-i}(s_i), \\ Z(s_i) & \text{si } |Z(s_i) - Z_{-i}^{mp}(s_i)| \leq c\sigma_{-i}(s_i), \end{cases}$$

où c est un seuil constant. Des valeurs comprises entre 1.5 et 2.5 ont montré qu'elles fournissaient des résultats satisfaisants.

V : Terminer en appliquant le krigage ordinaire aux observations éditées.

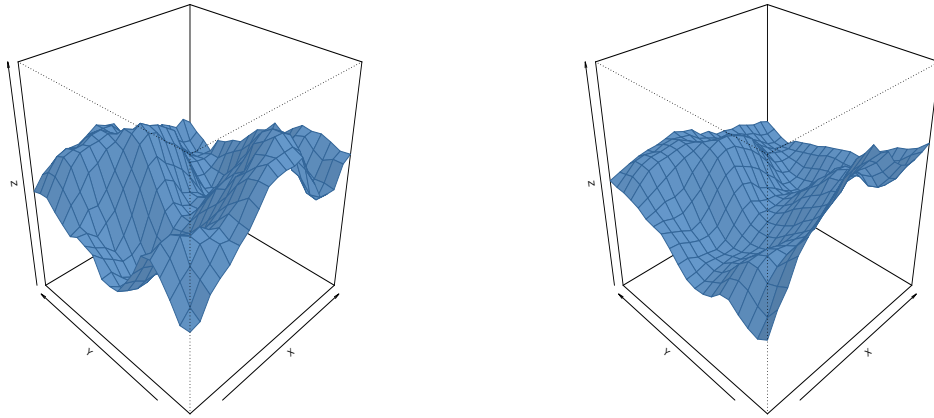


Figure 1.5. Représentations des surfaces obtenues après avoir appliqué le krigage robuste aux deux surfaces de la Figure 1.2. Le seuil constant pour l'édition des observations a été fixé à $c = 2$. Cette édition des observations justement est responsable de l'obtention de surface plus lisse que celles obtenues par krigage ordinaire et représentées dans la Figure 1.4.

La Figure 1.5 présente des surfaces obtenues après application du krigage robuste aux données. La médiane pondérée utilisée pour le krigage robuste, donnée par (1.19) est obtenue comme présentée dans Cressie (1993, p.147-148). On supposera ici que la séquence d'observations $\{Z(s_i) : i = 1, \dots, n, i \neq j\}$ est ordonnée de la plus petite à la plus grande des valeurs. On définit $\omega_{ji} = \lambda_{ji} / \sum_{t \neq j} \lambda_{ti}$ si $i \neq j$ et $\omega_{ji} \equiv 0$ si $i = j$. Les poids étant ceux du krigage ordinaire, ils ne sont pas forcément positifs. On cherche ensuite les intervalles

$$[Z(s_{l+1}), Z(s_{l+k+1})], \quad k \geq 0,$$

tel que

$$\sum_{i=1}^l \omega_{ji} \neq \sum_{i=1}^{l+1} \omega_{ji} = \dots = \sum_{i=1}^{l+k} \omega_{ji} = 1/2 \neq \sum_{i=1}^{l+k+1} \omega_{ji}.$$

La condition $k = 0$ correspond à la situation

$$\sum_{i=1}^l \omega_{ji} < 1/2 < \sum_{i=1}^{l+1} \omega_{ji} \quad \text{ou} \quad \sum_{i=1}^l \omega_{ji} > 1/2 > \sum_{i=1}^{l+1} \omega_{ji} \quad .$$

Tous ces points ou intervalles peuvent être dénommés médianes pondérées. Pour obtenir une solution unique, on considère leur médiane ou la médiane de leurs centres.

Ce chapitre, maintenant terminé, se veut une présentation relativement rapide de la géostatistique. Il représente à mes yeux un résumé succinct, mais consistant, du sujet. Le prochain chapitre présente une nouvelle manière de modéliser un processus présentant une hétérogénéité spatiale au niveau de l'échelle et les méthodes développées pour traiter un tel processus.

Modèle à erreurs substitutives

Une modélisation classique lors de l'étude d'un processus aléatoire spatial est de supposer que celui-ci possède une constitution additive. Autrement dit, le processus étudié $Y(\cdot)$ peut s'écrire comme

$$Y(s) = T(s) + Z(s), \quad s \in \mathcal{D}. \quad (2.1)$$

Dans cette décomposition, le premier processus, noté $T(\cdot)$, représente une tendance ; ce processus est sensé absorber la variation à basse résolution présente dans le jeu de données. Le second processus $Z(\cdot)$ est généralement supposé être un processus stationnaire au second ordre comme présenté dans la Proposition 1.3. Ce dernier peut encore éventuellement être décomposé en trois parties distinctes comme suggéré par Cressie (1993, p.112) :

$$Z(s) = W(s) + \eta(s) + \varepsilon(s), \quad s \in \mathcal{D}, \quad (2.2)$$

où

$W(\cdot)$ est un processus centré L_2 -continu. Autrement dit, son variogramme est continu (en particulier à l'origine) et sa portée est supérieure à la distance minimale séparant deux sites d'observations.

$\eta(\cdot)$ est un processus centré intrinsèquement stationnaire indépendant de $W(\cdot)$ et dont la portée est finie et inférieure à la distance minimale séparant deux sites d'observations.

$\varepsilon(\cdot)$ est un processus bruit-blanc indépendant des deux processus précédents.

Ce qu'il est important de retenir est que quelque soit le degré de décomposition choisi, cette dernière n'est pas unique. Basé sur le même jeu de données, deux scientifiques peuvent obtenir des ajustements différents pour les différents éléments de leur décomposition respective. En particulier, il faut remarquer que le comportement du processus $\eta(\cdot)$ est par définition inidentifiable sans informations en sus des n observations et de leurs localisations.

Dans la pratique, le processus de tendance est en premier lieu ajusté. Dans un deuxième temps, sur la base des résidus obtenus, le processus stationnaire au second ordre est ajusté par krigeage. Dans ce chapitre, l'attention sera focalisée sur l'ajustement du processus $Z(\cdot)$. Puisqu'assumer la gaussiannité des données est un paradigme pour le krigeage ordinaire (même si cela nécessite une transformation des données), la question s'est naturellement posée quant au traitement de valeurs aberrantes ou plus généralement de la présence d'une hétérogénéité conséquente dans le jeu de données. Une idée est de supposer que le processus $W(\cdot)$ est bien gaussien, mais que le processus d'"erreur" $\eta(\cdot) + \varepsilon(\cdot)$ ne l'est pas. En lieu et place, on suppose que sa distribution est donnée par

$$\begin{cases} \mathcal{N}(0, \sigma^2) & \text{avec probabilité } 1 - \epsilon, \\ \mathcal{L} & \text{avec probabilité } \epsilon, \end{cases} \quad (2.3)$$

où $0 < \epsilon \ll 1$ et \mathcal{L} représente une distribution centrée à variance finie et queues lourdes. Cela signifie que la majorité des données sont toujours modélisées par une loi gaussienne, mais qu'une petite partie d'entre elles ne respectent plus cette hypothèse et proviennent d'une loi dont le second moment est nettement plus élevé. Le variogramme s'écrit

$$2\gamma_Z(h) = 2\gamma_W(h) + 2 \left\{ (1 - \epsilon)\sigma^2 + \epsilon \int t^2 \mathcal{L}(dt) \right\}.$$

Puisque les observations provenant de la loi à queues lourdes sont considérées comme des erreurs, il est clair que ce n'est pas le variogramme ci-dessus qui doit être estimé. Le processus d'intérêt s'écrit $M(s) = \mu + W(s) + \delta(s)$, où $\delta(s) \sim \mathcal{N}(0, \sigma^2)$. Par conséquent, le variogramme à ajuster est

$$2\gamma_Z(h) = 2\gamma_W(h) + 2\sigma^2.$$

Cressie and Hawkins (1980a) et Hawkins and Cressie (1980) ont montré comment estimer ce variogramme. Par opposition au modèle développé ci-après, le modèle muni du processus d'erreur fourni par (2.2) et (2.3) est appelé **modèle à erreurs additives**.

La motivation principale de ce chapitre est l'idée que certaines des observations obtenues ne reflètent pas la réalité du processus d'intérêt, mais une contamination de celui-ci.

Cette idée peut servir de modélisation à différentes situations. Un exemple pourrait être un disfonctionnement des appareils de mesure. Un autre pourrait être la présence naturelle de valeurs éloignées de manière surprenante des valeurs attendues. Un exemple concret est la découverte retardée du trou dans la couche d’ozone au-dessus de l’Antartique, due principalement à un traitement informatique des enregistrements qui avait pour tâche de classer comme erreurs les chutes brutales de concentration. Le but de la méthode est donc de produire un estimateur de krigeage qui ne se laisse pas influencer par de telles valeurs et qui surtout ne rejette pas à la base des valeurs surprenantes.

Il est supposé dans ce chapitre que le processus spatial étudié $Z(\cdot)$ est un processus gaussien stationnaire au second ordre. Comme dans le modèle additive présenté dans l’introduction de ce chapitre, on va supposer que l’on n’observe pas directement le processus d’intérêt, mais une version “polluée” de celui-ci. Par contre, l’idée est que la contamination quand elle apparaît éclipse totalement l’observation, au lieu de s’y ajouter. D’où le nom du modèle : **modèle à erreurs substitutives**. Formellement, on écrira :

$$X(s) = (1 - B(s))Z(s) + B(s)O(s), \quad s \in \mathcal{D}, \quad (2.4)$$

où

$Z(\cdot)$ est le processus d’intérêt gaussien et stationnaire au second ordre, d’espérance $E(Z(s)) = \mu$, $\forall s \in \mathcal{D}$ et de variance $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq n}$ tel que $\sigma_{ij} = \sigma^2$, $\forall i = j$;

$O(\cdot)$ est le processus de contamination. Il possède une distribution \mathcal{L} centrée autour de μ , de variance finie et à queues lourdes. Il est stationnaire au second ordre et est supposé indépendant de $Z(\cdot)$;

$B(\cdot)$ est un processus binaire intrinsèquement stationnaire, indépendant de $Z(\cdot)$ et de $O(\cdot)$. Pour chaque site d’observation s , il indique si la valeur observée provient du processus d’intérêt ou du processus de contamination. On notera $B(\cdot) \sim \text{Bernoulli}(\epsilon)$. Le paramètre ϵ représente le taux de contamination du modèle. Dans un premier temps, aucune dépendance spatiale pour ce processus n’est supposée. De même, la probabilité d’exhiber une valeur contaminée ne dépend pas du site d’observation.

Le chapitre se décompose comme suit. Dans un premier temps le nouveau modèle est présenté. L’estimateur de krigeage adéquat associé est dérivé tout d’abord sous l’hypothèse de normalité de l’erreur. Des simulations illustrent les qualités de l’estimateur présenté. Par la suite, une application de la méthode lors d’une participation à un concours d’interpolation spatiale est présenté. Finalement, une généralisation du modèle est proposée et des simulations effectuées. Mais, tout d’abord, le variogramme associé à la modélisation est dérivé.

2.1 Variogramme associé au modèle

Dans cette section, le variogramme théorique du processus $X(\cdot)$ est dérivé en fonction des variogrammes des processus sous-jacents : $Z(\cdot)$, $O(\cdot)$ et $B(\cdot)$. Pour ce faire, on définit encore le processus conditionné

$$V(s) = X(s) \mid (B(s) = b(s)), \quad s \in \mathcal{D}. \quad (2.5)$$

Théorème 2.1 Le variogramme du processus $X(\cdot)$ donné par (2.4) s'écrit :

$$\begin{aligned} 2\gamma_X(s_1 - s_2) &= \left(1 - \epsilon - \gamma_B(s_1 - s_2)\right) \cdot 2\gamma_Z(s_1 - s_2) \\ &\quad + \left(\epsilon - \gamma_B(s_1 - s_2)\right) \cdot 2\gamma_O(s_1 - s_2) \\ &\quad + 2\gamma_B(s_1 - s_2) \cdot 2\left(\frac{\sigma_Z^2 + \sigma_O^2}{2}\right), \quad \forall s_1, s_2 \in \mathcal{D}, \end{aligned} \quad (2.6)$$

où $\epsilon = E(B(s))$, $\sigma_Z^2 = \text{Var}(Z(s))$ et $\sigma_O^2 = \text{Var}(O(s))$ pour tout $s \in \mathcal{D}$.

Preuve 2.2 Si le variogramme de $X(\cdot)$ existe, on peut écrire

$$\begin{aligned} 2\gamma_X(s_1 - s_2) &= E\left(\text{Var}\left(X(s_1) - X(s_2) \mid B(s_1), B(s_2)\right)\right) \\ &\quad + \text{Var}\left(E\left(X(s_1) - X(s_2) \mid B(s_1), B(s_2)\right)\right), \\ &= E\left(\text{Var}\left(V(s_1) - V(s_2)\right)\right) + \text{Var}\left(E\left(V(s_1) - V(s_2)\right)\right) \\ &= E\left(2\gamma_V(s_1 - s_2)\right), \end{aligned}$$

puisque $E(V(s)) = \mu$, $\forall s \in \mathcal{D}$. Le variogramme du processus $V(\cdot)$ s'écrit

$$\begin{aligned} 2\gamma_V(s_1 - s_2) &= \text{Var}\left(X(s_1) - X(s_2) \mid B(s_1), B(s_2)\right) \\ &= \text{Var}\left(\left(Z(s_1) - Z(s_2)\right) + b(s_1)\left(O(s_1) - Z(s_1)\right) + b(s_2)\left(Z(s_2) - O(s_2)\right)\right) \\ &= \left(1 - (b(s_1) + b(s_2)) + b(s_1)b(s_2)\right) \cdot 2\gamma_Z(s_1 - s_2) + b(s_1)b(s_2) \cdot 2\gamma_O(s_1 - s_2) \\ &\quad + \left(b(s_1) - b(s_2)\right)^2 \cdot 2\left(\frac{\sigma_Z^2 + \sigma_O^2}{2}\right). \end{aligned}$$

Puisque $2\gamma_X(s_1 - s_2) = E_B(2\gamma_V(s_1 - s_2))$, il reste maintenant à calculer cette espérance. La distribution conjointe de $B(s_1)$ et $B(s_2)$ est présentée dans le tableau ci-après.

	$B(s_2) = 1$	$B(s_2) = 0$	
$B(s_1) = 1$	$\epsilon - \gamma_B(s_1 - s_2)$	$\gamma_B(s_1 - s_2)$	ϵ
$B(s_1) = 0$	$\gamma_B(s_1 - s_2)$	$1 - \epsilon - \gamma_B(s_1 - s_2)$	$1 - \epsilon$
	ϵ	$1 - \epsilon$	1

Tableau 2.1. Distribution bivar  e du processus de localisation de la contamination $B(\cdot)$.

Par cons  quent,

$$\begin{aligned}
E_B(2\gamma_V(s_1 - s_2)) &= \mathbf{P}(B(s_1) = 0, B(s_2) = 0) \cdot 2\gamma_Z(s_1 - s_2) \\
&\quad + \mathbf{P}(B(s_1) = 1, B(s_2) = 1) \cdot 2\gamma_O(s_1 - s_2) \\
&\quad + \mathbf{P}(B(s_1) \neq B(s_2)) \cdot 2\left(\frac{\sigma_Z^2 + \sigma_O^2}{2}\right) \\
&= \left(1 - \epsilon - \gamma_B(s_1 - s_2)\right) \cdot 2\gamma_Z(s_1 - s_2) \\
&\quad + \left(\epsilon - \gamma_B(s_1 - s_2)\right) \cdot 2\gamma_O(s_1 - s_2) + 2\gamma_B(s_1 - s_2) \cdot 2\left(\frac{\sigma_Z^2 + \sigma_O^2}{2}\right).
\end{aligned}$$

◆

On en conclut que le processus $X(\cdot)$ est un processus intrins  quement stationnaire. Dans la section suivante, un premier cas particulier est   tudi  . Il s'agit de la situation o   les variogrammes des processus de contamination et de localisation de la contamination, $O(\cdot)$ et $B(\cdot)$ sont en fait des variogrammes de bruits blancs. Autrement dit, il est suppos   qu'il n'y a pas de d  pendance spatiale, ni dans le processus de localisation, ni dans le processus d'intensit   de la contamination. L'estimateur optimal au sens du carr   moyen de l'erreur sans biais est d  riv  .

2.2 Contamination gaussienne

Tukey (1960) a montr   que si on consid  re le simple mod  le

$$Z(s) = \mu + \varepsilon(s), \quad s \in \mathcal{D},$$

o  

$$\varepsilon(s) \stackrel{\text{iid}}{\sim} \begin{cases} \mathcal{N}(0, \sigma^2) & \text{avec probabilit   } 1 - \epsilon, \\ \mathcal{N}(0, k^2 \sigma^2) & \text{avec probabilit   } \epsilon, \end{cases}$$

le choix des paramètres $\epsilon = 0.05$ et $k^2 = 9$ suffit amplement à détruire l'optimalité de la moyenne usuelle \bar{Z} pour l'estimation du paramètre μ . Même si le problème est différent, présence d'une dépendance spatiale, une caractérisation similaire du processus de contamination bien que simpliste suffit à illustrer le problème et la nécessité du développement d'un nouvel estimateur.

Pour illustration, on va tout d'abord considérer le cas unidimensionnel $d = 1$ et $\mathcal{D} \subset \mathbb{R}$. Le but est de prédire la valeur du processus d'intérêt $Z(\cdot)$ au point s_2 , basé sur l'observation du processus $X(\cdot)$ au point s_1 . Les variables $Z(s_1)$ et $Z(s_2)$ possèdent une distribution bivariable normale d'espérance $\mu = (\mu_1, \mu_2)^t$ et de matrice de covariance $\Sigma = (\sigma_{ij})_{1 \leq i, j \leq 2}$. La variable $O(s_1)$, indépendante des deux variables $Z(s_1)$ et $Z(s_2)$, possède une distribution normale d'espérance μ_1 et de variance $\sigma_O^2 = k^2 \cdot \sigma_{11}$. Finalement, $B(s_1) \sim \text{Bernoulli}(\epsilon)$ est indépendante de toutes les autres variables et le modèle s'écrit

$$X(s_1) = (1 - B(s_1))Z(s_1) + B(s_1)O(s_1).$$

Dans ce cas, la densité conjointe de $X_1 = X(s_1)$ et de $Z_2 = Z(s_2)$ s'écrit

$$f_{X_1, Z_2} = (1 - \epsilon)f_{Z_1, Z_2} + \epsilon f_{O_1, Z_2}.$$

L'espérance conditionnelle est donnée par

$$\begin{aligned} \hat{Z}(s_2) &= E(Z(s_2) | X(s_1)) \\ &= \int z_2 f_{X_1, Z_2}(x_1, z_2) dz_2 \Bigg/ \int f_{X_1, Z_2}(x_1, z_2) dz_2 \\ &= \frac{(1 - \epsilon)f_{Z_1}(x_1)E(Z_2 | X_1 = x_1, B_1 = 0) + \epsilon f_{O_1}(x_1)E(Z_2 | X_1 = x_1, B_1 = 1)}{(1 - \epsilon)f_{Z_1}(x_1) + \epsilon f_{O_1}(x_1)} \\ &= (1 - \omega(x_1)) \frac{\mu_2 + \sigma_{21}(x_1 - \mu_1)}{\sqrt{\sigma_{11}}} + \omega(x_1)\mu_2, \end{aligned}$$

où $\omega(x) = \epsilon f_{O_1}(x) / ((1 - \epsilon)f_{Z_1}(x) + \epsilon f_{O_1}(x))$, représentée dans la Figure 2.1 à gauche, correspond au rapport de vraisemblance.

L'estimateur optimal est donc une moyenne pondérée entre l'estimateur classique correspondant à l'absence de contamination, $E(Z_2 | X_1)$, et l'espérance inconditionnelle du processus, $E(Z_2)$, lorsque la valeur observée ne véhicule aucune information. Les poids respectifs correspondent à la fonction de vraisemblance associée aux données conditionnée sur la connaissance du scénario de contamination : $f_{X_1, B_1}(x_1, 0) = f_{X_1 | B_1}(x_1) \mathbf{P}(B_1 = 0)$.

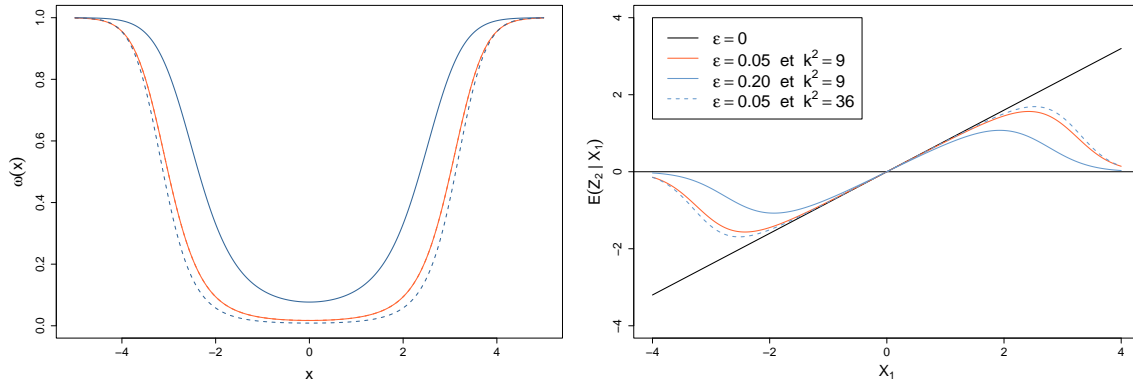


Figure 2.1. A gauche, la représentation de la fonction $\omega(x)$ pour un cas centré. La courbe orange correspond à $(\epsilon = 0.05, k^2 = 9)$. Les deux autres courbes sont obtenues en faisant varier un paramètre ou l'autre. La courbe en trait plein correspond à $\epsilon = 0.2$ et celle en traitillé à $k^2 = 36$. Cette courbe représente le poids accordé à l'espérance inconditionnelle dans la prévision. On remarque que varier la probabilité de contamination dans ce modèle a plus d'influence que le fait de varier le paramètre multiplicatif k^2 . Dans le graphique de droite, les courbes de régression pour les différentes combinaisons des paramètres. La droite diagonale correspond à $\epsilon = 0.00$. Autrement dit, il s'agit de l'estimateur classique dans le cas d'une loi bivariée normale. La corrélation entre $Z(s_1)$ et $Z(s_2)$ a été fixée à $\sigma_{12} = 0.8$. Plus on augmente le taux de contamination, plus le prédicteur se rabat rapidement sur l'espérance inconditionnelle. A l'inverse, si l'on augmente le paramètre k^2 , l'estimateur sachant qu'il doit décider entre deux lois d'autant plus distinctes, s'attarde d'autant plus longtemps près de l'estimateur classique avant de se rabattre sur le centre de la distribution de $Z(s_2)$.

2.2.1 Le modèle

Pour rappel, on suppose dans cette section qu'il n'existe pas de dépendance spatiale pour les processus $O(\cdot)$ et $B(\cdot)$. Si l'on suppose qu'il y a n sites d'observations, on peut écrire que

$$Z = (Z(s_1), \dots, Z(s_n))^t \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma),$$

où $\boldsymbol{\mu} = \mu \cdot \mathbf{1}$ et $\Sigma = (\sigma_{ij})$ telle que $\sigma_{ii} = \mathcal{C}(0) = \sigma^2$, $\forall i = 1, \dots, n$. De la même manière,

$$O = (O(s_1), \dots, O(s_n))^t \sim \mathcal{N}(\boldsymbol{\mu}, k^2 \cdot \sigma^2 \cdot \mathbf{I}_n),$$

où \mathbf{I}_n est la matrice identité de dimension n et

$$B = (B(s_1), \dots, B(s_n))^t \sim \text{Bernoulli}(\epsilon),$$

où $B(s_i) \perp B(s_j)$, $\forall i \neq j$. Ces trois vecteurs sont indépendants entre eux, composante par composante.

Le variogramme donné par (2.6) s'écrit, dans ce cas comme

$$2\gamma_X(s_1 - s_2) = \begin{cases} 0, & \text{si } s_1 - s_2 = 0, \\ (1 - \epsilon)^2 \cdot 2\gamma_Z(s_1 - s_2) + \epsilon((1 - \epsilon) + \epsilon \cdot k^2) \cdot 2\sigma^2, & \text{si } s_1 - s_2 \neq 0. \end{cases}$$

Le but est maintenant de déterminer l'écriture de l'estimateur optimal au sens du carré moyen de l'erreur. On cherche donc un estimateur tel que

$$E \left((\widehat{Z}(s_0) - Z(s_0))^2 \right) = \min!.$$

Le processus $Z(\cdot)$ comme le processus $O(\cdot)$ étant gaussiens et indépendants l'un de l'autre, le processus conditionné $V(\cdot)$ est également gaussien. En fait,

$$V = (V(s_1), \dots, V(s_n))^t \sim \mathcal{N}(\boldsymbol{\mu}, \Omega_b),$$

où

$$\Omega_b = (I_n - A_b)\Sigma(I_n - A_b) + k^2\sigma^2 A_b$$

et $A_b = \text{diag}(b) \in \{0, 1\}^{n \times n}$ avec $b = (b(s_1), \dots, b(s_n))^t$. La covariance entre les vecteurs aléatoires Z et V vaut

$$\text{Cov}(Z, V) = \text{Cov}(Z, (I_n - A_b)Z + A_b O) = \Sigma(I_n - A_b).$$

En particulier, on a que $\text{Cov}(Z(s_0), V) = \nu^t(I_n - A_b)$ où $\nu^t = \text{Cov}(Z(s_0), Z)$. Puisque le processus $Z(\cdot)$ est supposé stationnaire au second ordre, le vecteur ν ne dépend que de la distance séparant les différents sites d'observation, via le variogramme de $Z(\cdot)$. D'autre part, la matrice Ω_b étant symétrique, sa matrice inverse généralisée est définie et s'écrit $\Omega_b^- = \Phi \Delta^- \Phi$, où les colonnes de Φ représentent les vecteurs propres orthogonaux correspondant aux valeurs propres non-nulles contenues dans la matrice diagonale Δ . Cette matrice vérifie les propriétés

$$\begin{aligned} \Omega_b \cdot \Omega_b^- \cdot \Omega_b &= \Omega_b, \\ \Omega_b^- \cdot \Omega_b \cdot \Omega_b^- &= \Omega_b^-. \end{aligned}$$

Les matrices A_b et Ω_b dépendent directement du scénario de contamination b . Le nombre n représentant le nombre de sites d'observation, on obtient un total de 2^n scénarios de contamination distincts. On note \mathcal{B} , l'ensemble de tous les scénarios possibles de cardinalité 2^n . On peut énoncer le résultat suivant.

Théorème 2.3 L'estimateur sans biais optimal au sens du carré moyen de l'erreur est donné par

$$\widehat{Z}(s_0) = \sum_{b \in \mathcal{B} = \{0,1\}^n} \omega_b \left(\nu^t (\mathbf{I}_n - A_b) + \frac{1 - \nu^t (\mathbf{I}_n - A_b) \Omega_b^- \mathbf{1}}{\mathbf{1}^t \Omega_b^- \mathbf{1}} \cdot \mathbf{1}^t \right) \Omega_b^- X, \quad (2.7)$$

où

$$\omega_b = \frac{\mathbf{P}(B = b) \mathbf{P}(X = x | B = b)}{\mathbf{P}(X = x)} = \frac{\epsilon^{n_1} \cdot (1 - \epsilon)^{n - n_1} \mathbf{P}(X = x | B = b)}{\mathbf{P}(X = x)}, \quad (2.8)$$

et $n_1 = \#\{b(s_i) = 1 : i = 1, \dots, n\}$ représente le nombre de sites contaminés dans le scénario b .

Preuve 2.4 Puisque la fonction de perte est l'erreur carrée, l'estimateur optimal est fourni par l'espérance conditionnelle $\mathbf{E}(Z(s_0) | X)$. On peut écrire

$$\begin{aligned} \mathbf{E}(Z(s_0) | X) &= \int z_0 f(z_0 | x) \, dz_0 \\ &= \int \int z_0 f(z_0 | x, b) f(b | x) \, db \, dz_0 \end{aligned}$$

car

$$\begin{aligned} f(z_0 | x) &= \frac{f(z_0, x)}{f(x)} \\ &= \int \frac{f(z_0, x, b)}{f(x)} \, db = \int \frac{f(z_0 | x, b) f(x, b)}{f(x)} \, db \\ &= \int \frac{f(z_0 | x, b) f(x | b) f(b)}{f(x)} \, db = \int f(z_0 | x, b) f(b | x) \, db. \end{aligned}$$

Par conséquent,

$$\begin{aligned} \mathbf{E}(Z(s_0) | X) &= \int \int z_0 f(z_0 | x, b) \, dz_0 \frac{f(x | b) f(b)}{f(x)} \, db \\ &= \int \mathbf{E}(Z(s_0) | X, B) \frac{f(x | b) f(b)}{f(x)} \, db = \int \mathbf{E}(Z(s_0) | V) \frac{f(x | b) f(b)}{f(x)} \, db. \end{aligned}$$

L'ensemble \mathcal{B} étant fini dénombrable, on peut réécrire cette dernière équation comme

$$\widehat{Z}(s_0) = \sum_{b \in \mathcal{B} = \{0,1\}^n} \omega_b \mathbf{E}(Z(s_0) | V),$$

où $\omega_b = \mathbf{P}(B = b) \mathbf{P}(X = x | B = b) / \mathbf{P}(X = x)$, $\forall b \in \mathcal{B}$. Il reste donc à calculer l'espérance conditionnelle $\mathbf{E}(Z(s_0) | V)$. Puisque $V(\cdot)$ est gaussien, cette espérance est

confondue avec l'estimateur optimal linéaire. Autrement dit, elle est confondue avec la solution du krigeage ordinaire appliqué aux données conditionnées V . On peut donc réécrire le problème comme

$$\widehat{Z}(s_0) = \sum_{b \in \mathcal{B}} \omega_b \widehat{Z}_b(s_0) = \sum_{b \in \mathcal{B}} \omega_b \alpha_b^t V,$$

où $\sum_{b \in \mathcal{B}} \omega_b = 1$ et $\sum_{i=1}^n \alpha_{bi} = 1$ de sorte que $\widehat{Z}_b(s_0)$ représente l'estimateur linéaire optimal sans biais si on connaît le scénario de contamination. Dans un premier temps, on cherche l'écriture de α_b . Le problème devient

$$\min_{\alpha_b, m} \left(\mathbb{E} \left((Z(s_0) - \widehat{Z}_b(s_0))^2 \right) - 2m(\alpha_b^t \mathbf{1} - 1) \right), \quad (2.9)$$

où m représente le multiplicateur de Lagrange associé à cette minimisation sous contrainte. Puisque

$$\begin{aligned} (Z(s_0) - \widehat{Z}_b(s_0))^2 &= ((Z(s_0) - \mu) - \alpha_b^t (V - \mu \mathbf{1}))^2, \\ &= (Z(s_0) - \mu)^2 + (\alpha_b^t (V - \mu \mathbf{1}))^2 - 2\alpha_b^t (Z(s_0) - \mu)(V - \mu \mathbf{1}), \end{aligned}$$

en utilisant le fait que les poids doivent sommer à 1, on peut réécrire l'espérance présente dans (2.9) comme

$$\begin{aligned} \mathbb{E} \left((Z(s_0) - \widehat{Z}_b(s_0))^2 \right) &= \text{Var}(Z(s_0)) + \alpha_b^t \text{Var}(V) \alpha_b - 2\alpha_b \text{Cov}(V, Z(s_0)) \\ &= \sigma^2 + \alpha_b^t \Omega_b \alpha_b - 2\alpha_b^t (\mathbf{I}_n - A_b) \nu. \end{aligned}$$

En dérivant la fonction à minimiser donnée par (2.9) et en résolvant le système d'équation obtenu en posant les dérivées égales à zéro, on obtient le système de solutions suivant :

$$\begin{cases} m = \frac{1 - \nu^t (\mathbf{I}_n - A_b) \Omega_b^- \mathbf{1}}{\mathbf{1}^t \Omega_b^- \mathbf{1}}, \\ \alpha_b^t = \left(\nu^t (\mathbf{I}_n - A_b) + m \cdot \mathbf{1}^t \right) \cdot \Omega_b^- . \end{cases} \quad (2.10)$$

◆

Cet estimateur représente l'estimateur optimal au sens du carré moyen de l'erreur. Cependant son évaluation nécessite un temps de calcul exponentiel, même pour une valeur raisonnablement grande de n . L'idée est donc de trouver un sous-ensemble S de scénarios contenus dans \mathcal{B} tel que le temps de calcul devienne raisonnable et la perte d'optimalité induite minimale. Pour déterminer ce sous-ensemble, il apparaît intuitif de souhaiter que

les scénarios les plus probables s'y retrouvent. Par conséquent, l'estimateur proposé est le suivant.

Définition 2.5 Un estimateur simplifié est donné par

$$\widehat{Z}_S(s_0) = \sum_{b \in S} \tilde{\omega}_b \left(\nu^t (\mathbf{I}_n - A_b) + \frac{1 - \nu^t (\mathbf{I}_n - A_b) \Omega_b^- \mathbf{1}}{\mathbf{1}^t \Omega_b^- \mathbf{1}} \cdot \mathbf{1}^t \right) \Omega_b^- X, \quad (2.11)$$

où les poids $\tilde{\omega}_b$ sont proportionnels à la fonction de vraisemblance

$$\mathbf{P}(B = b) \mathbf{P}(X = x | B = b).$$

et sont tels que

$$1 \approx \sum_{b \in S} \tilde{\omega}_b \gg \sum_{b \notin S} \tilde{\omega}_b \approx 0.$$

Une fois l'ensemble S défini, les poids associés aux scénarios retenus sont normés pour assurer que l'estimateur obtenu soit toujours sans biais.

En effet, en considérant l'écriture des poids $\tilde{\omega}_b$, on est en droit de s'attendre à ce qu'une minorité d'entre eux seulement soient non négligeables (voir Figure 2.8). D'où l'idée de trouver un sous-ensemble de scénarios tel que l'estimateur soit calculable en un temps raisonnable. Les Figures 2.2 et 2.3 sont là pour illustrer la motivation à développer une telle méthode. On peut encore remarquer que l'estimateur présenté ici est bien l'analogue multidimensionnel de celui présenté au début de la Section 2.2. En effet, les poids donné par l'équation (2.8) peuvent être réécrits comme

$$\begin{aligned} \omega_b &= \frac{\mathbf{P}(B = b) \mathbf{P}(X = x | B = b)}{\mathbf{P}(X = x)} \\ &= \frac{\mathbf{P}(B = b) \mathbf{P}(X = x | B = b)}{\int \mathbf{P}(B = b) \mathbf{P}(X = x | B = b) db}. \end{aligned}$$

Cette dernière équation montre bien que chaque poids ω_b correspond à la vraisemblance associée au scénario b normalisée par la vraisemblance inconditionnelle associée aux observations X .

Ci-après est présentée la méthode de détermination du sous-ensemble S qui est sensé contenir les scénarios de contamination les plus probables ; autrement dit, ceux dont les poids $\tilde{\omega}_b$ sont les plus élevés. La démarche se compose de deux étapes.

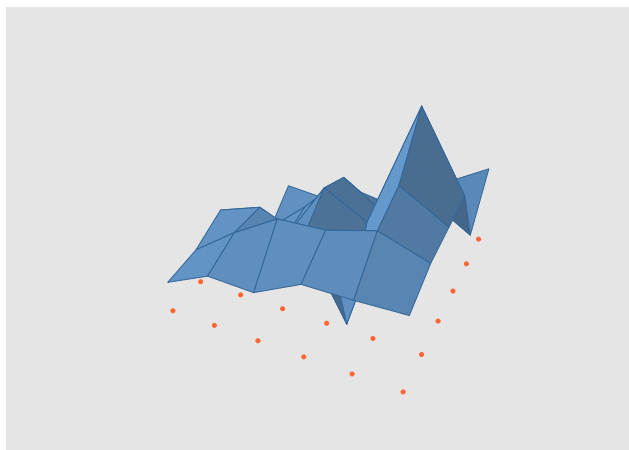


Figure 2.2. Une réalisation d'un modèle à erreurs substitutives simulée sur une grille régulière de 6×6 localisations. Le variogramme du processus d'intérêt est sphérique de paramètre $\theta = (0.05, 0.95, 4)^t$. Le taux de contamination vaut $\epsilon = 0.05$ et $k^2 = 9$. Deux sites sont effectivement contaminés (valeurs : -5.9 et 5.4).

Etape 1 : Identifier un scénario de contamination vraisemblable, afin qu'il serve de scénario initial pour l'étape suivante.

Etape 2 : Déterminer un sous-ensemble de scénarios correspondant à un voisinage du scénario initial sur l'hypercube $\{0, 1\}^n$.

2.2.2 Etape 1 : Détermination d'un scénario de contamination initial

La méthode utilisée est basée sur la technique de détection de valeurs aberrantes proposée par Cerioli and Riani (1999). Les valeurs aberrantes spatiales sont généralement définies comme des valeurs extrêmes comparées aux valeurs voisines. Le problème majeur étant qu'elles peuvent se présenter isolées, groupées ou dispersées. Les méthodes classiques, comme celle suggérée par Cressie (1986), basées sur les résidus de prédiction standardisés fonctionnent bien si les valeurs aberrantes sont isolées et uniques. En cas de valeurs aberrantes multiples, ces statistiques de détection/effacement de cas uniques subissent des effets secondaires tel le masquage des autres valeurs sujettes à caution. Justement, Cerioli and Riani (1999) défendent le fait que la méthode appelée *forward search* ne subit pas ces effets similaires et possède la même efficacité quelque soit la distribution dans l'espace des valeurs aberrantes. L'idée principale de la méthode est d'ordonner les observations de la plus en accord avec le processus d'intérêt à celle le plus en désaccord. La méthode utilisée dans ce travail est présentée ci-après. Elle est constituée de quatre points.

valide, les estimations servent ensuite à ajuster un variogramme paramétrique valide par la méthode des moindres carrés. Une illustration de ce premier point est fourni par la Figure 2.4.

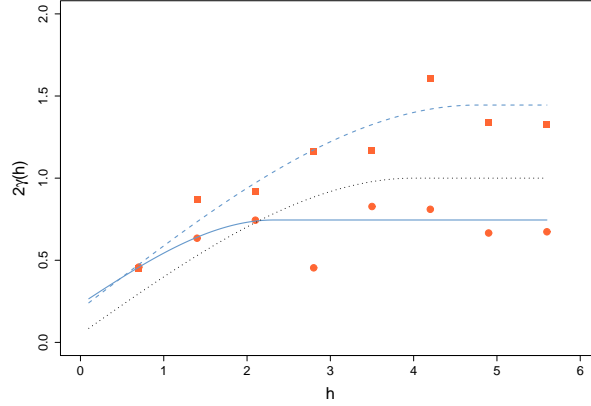


Figure 2.4. Ajustement du variogramme associé aux données présentées dans la Figure 2.2. Les carrés oranges correspondent à l'estimateur proposé par Cressie and Hawkins (1980b) ; les ronds oranges à celui proposé par Genton (1998a). Les courbes sont obtenues en ajustant un variogramme sphérique aux estimations. La courbe noire correspond au vrai variogramme.

Point II : Détermination d'un sous-ensemble initial

La recherche commence par la détermination d'un sous-ensemble de $p \ll n$ localisations spatiales. Ce sous-ensemble est supposé contenir uniquement des localisations où aucune contamination ne s'est produite. On choisit le sous-ensemble qui satisfait au critère de moindre médiane des carrés des résidus standardisés. Si on dénote par $\hat{Z}_{i(S_l^{(p)})}$ la prédiction au site s_i basée sur les observations obtenues aux sites contenus dans $S_l^{(p)} = \{s_{i_1^l}, \dots, s_{i_p^l}\}$, $1 \leq i_1^l, \dots, i_p^l \leq n$, et par

$$\tilde{e}_{i(S_l^{(p)})} = \frac{Z(s_i) - \hat{Z}_{i(S_l^{(p)})}}{\sigma_{i(S_l^{(p)})}},$$

l'erreur de prédiction standardisée associée, le sous-ensemble initial $S_\star^{(p)}$ est choisi tel que

$$\tilde{e}_{[med](S_\star^{(p)})}^2 = \min_l e_{[med](S_l^{(p)})}^2,$$

où $\tilde{e}_{[k](S_l^{(p)})}^2$ est la $k^{\text{ème}}$ statistique d'ordre de la série $\tilde{e}_{i(S_l^{(p)})}^2$, $i = 1, \dots, n$ et $med = p + \lfloor (n - p)/2 \rfloor$. Ici, p a été choisi aussi petit que possible ; autrement dit, $p = 2$. L'idée est donc de sélectionner le p -tuple de sites tel que la statistique d'ordre médiane des carrés des erreurs de prédiction standardisées basées sur ces sites soit minimale (*cf.* Figure 2.5).

C'est entre autre la partie la plus coûteuse en temps de calcul de la méthode, puisque il y a typiquement $\binom{n}{p}$ p -tuples possibles. On peut réduire ce temps de calcul en choisissant $p > 2$ sites différemment. Par exemple, en variant de manière aléatoire un ensemble de départ de taille $p > 2$. Même si l'assurance que cet ensemble de départ ne contienne pas de sites contaminés n'est pas essentielle, la chance d'en inclure croît évidemment.

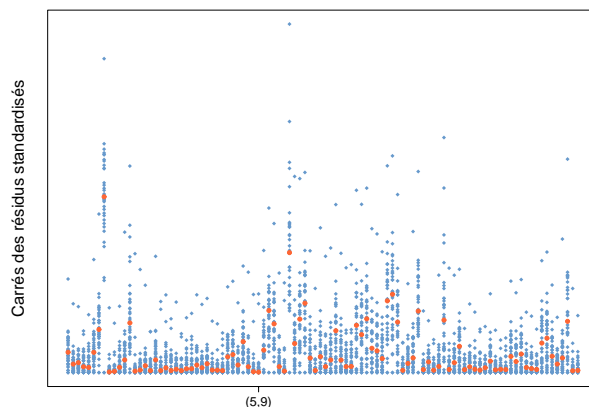


Figure 2.5. Représentation des carrés des résidus de prédiction standardisés en fonction des sous-ensembles de sites choisis pour établir la prédiction. Seulement une centaine de sous-ensembles (parmi un total de 2 parmi 36 possibilités) sont représentés. En orange, la médiane de chaque échantillon. Le couple composé des sites 5 et 9 minimise la valeur de cette médiane et est par conséquent choisi par l'algorithme pour commencer l'ordonnancement des données.

Point III : Progression

Etant donné un sous-ensemble de taille $m \geq p$, noté $S_{\star}^{(m)}$, la recherche progresse en sélectionnant les $m+1$ localisations générant les plus petits carrés de résidus de prédiction standardisés parmi la série

$$\tilde{e}_{i(S_{\star}^{(m)})}^2, \quad 1 \leq i \leq n.$$

Des variations de la méthode existent, en particulier si le sous-ensemble initial $S_{\star}^{(p)}$ est défini aléatoirement plutôt que de la manière présentée ici.

Point IV : Ordonnancement des sites

Une fois l'étape précédente terminée, l'algorithme nous fournit un ordre des données (ou sites associés); le premier site inclut représentant le site le plus en accord avec le modèle et le dernier le moins en accord. En termes du modèle (2.4), cela signifie que les observations contaminées provenant du processus $O(\cdot)$ sont sensées apparaître en dernier dans cette classification. Il s'agit maintenant de déterminer où se situe la démarcation entre les deux groupes d'observations. Cerioli and Riani (1999) suggèrent de se baser

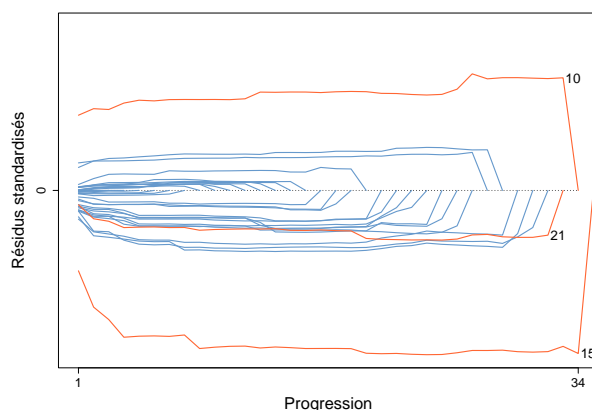


Figure 2.6. Evolution temporelle des résidus standardisés associés aux différents sites. Parti d'un sous-ensemble de $p = 2$ sites, l'algorithme progresse en ajoutant à chaque étape un site. Les trois derniers sites à rejoindre la liste sont les sites 21, 10 et 15 (dans l'ordre).

sur une observation minutieuse de différentes représentations graphiques de fonctions de statistiques d'ordre des résidus standardisés ou des écarts-types associés. Le choix de tel ou tel graphique dépend de la disposition des valeurs suspectes dans l'espace, à savoir si elles sont aléatoirement distribuées ou regroupées.

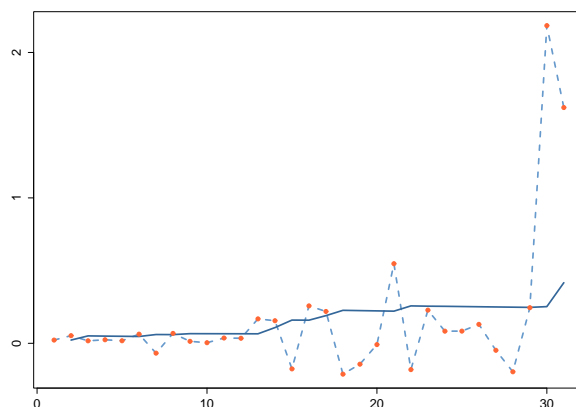


Figure 2.7. Illustration de la détermination du scénario initial fourni par la méthode *forward search* dans le cas du jeu de données représenté dans la Figure 2.2. Les points (en orange) correspondent aux valeurs de la suite définie par (2.13). La courbe en trait plein relie la suite des quantiles qui ont été définis dans la formule (2.14). On constate que les deux dernières valeurs de la suite différenciée dépassent le seuil fixé par la suite des quantiles. Cela signifie que les trois dernières observations ordonnées sont considérées comme observations contaminées.

Puisque l'établissement de ce scénario de contamination initial est dans la méthode proposée une première étape, il est souhaitable que cela soit fait de manière systématique

contrairement à ce qui est proposé. On définit la suite de valeurs

$$\tilde{e}_m^{(m+1)} = \sqrt{\tilde{e}_{[m+1](S_\star^{(m)})}^2}, \quad m = p+1, \dots, n-1, \quad (2.12)$$

où $\tilde{e}_{[m+1](S_\star^{(m)})}$ représente la $(m+1)^{\text{ème}}$ statistique d'ordre des résidus de prédiction standardisés résultant des prédictions effectuées sur la base des m premières observations ordonnées et contenues dans le sous-ensemble marqué $S_\star^{(m)}$. A partir de cette suite, on considère la suite différenciée

$$\check{e}_m^{(m+1)} = \tilde{e}_m^{(m+1)} - \tilde{e}_{m-1}^{(m)}, \quad m = p+1, \dots, n-1. \quad (2.13)$$

Basées sur cette suite, les règles de détermination de la contamination d'un site sont données ci-dessous. Une illustration de cette règle de détermination est fournie par la Figure 2.7.

Règle I : Le site qui correspond à l'élément $\check{e}_m^{(m+1)}$ est suspect si

$$\check{e}_m^{(m+1)} > q\hat{F}_{(\check{e}_p^{(p+1)}, \dots, \check{e}_{m-1}^{(m)})}(1 - \alpha), \quad (2.14)$$

pour $0 < \alpha < 1$ choisi. Autrement dit, on compare chacun des termes de cette suite à un quantile de la loi empirique associée aux termes le précédant dans la suite.

Règle II : Le premier site suspect (dans l'ordre obtenu à la section précédente) et tous les suivants sont déclarés contaminés.

2.2.3 Etape 2 : Détermination d'un sous-ensemble de scénarios

Maintenant que la première étape est terminée puisque un scénario de contamination initial a été obtenu, il s'agit de définir à partir de là un sous-ensemble de scénarios qualifiés de probables sur lesquels sera basée la prédiction. L'idée principale a été de formaliser ce problème comme celui de la recherche d'un voisinage d'un sommet sur l'hypercube unitaire $\{0, 1\}^n \subset \mathbb{R}^n$. En effet, chaque scénario de contamination $b(s)$ possède des composantes binaires et peut donc être vu comme un des sommets de l'hypercube. Le premier problème est le nombre de sommets existants. En effet, le total étant de 2^n , il n'est pas envisageable d'évaluer la vraisemblance associée à chaque sommet. C'est pourquoi, une politique de parcours des sommets doit être définie et la vraisemblance n'est évaluée que pour les sommets visités. Deux méthodes distinctes ont été envisagées pour parcourir les sommets de cet hypercube de manière efficace, mais seule la dernière a été retenue.

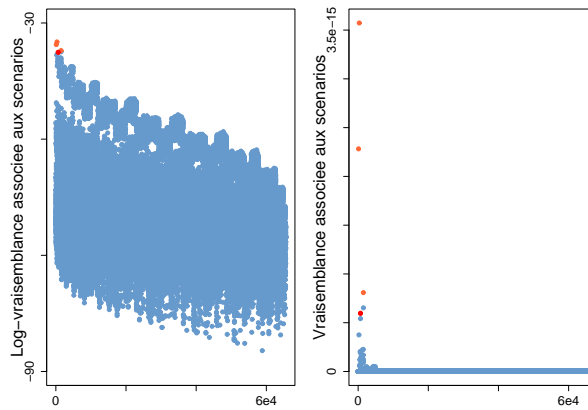


Figure 2.8. Représentation des poids associés à tous les sommets de l’hypercube plongé dans \mathbb{R}^{16} . Ces poids sont égaux à la vraisemblance associée à chaque sommet, calculée sur la base d’une réalisation d’un processus stochastique spatial sur une grille régulière de $4 \times 4 = 16$ points recouvrant le carré unitaire $[0, 1] \times [0, 1]$. En rouge, le poids associé au scénario initial trouvé par la méthode *forward search*. En orange les poids correspondant aux scénarios retenus pour la prédiction. On constate qu’il reste des poids conséquents non considérés, mais on remarque surtout que le poids le plus élevé est bel et bien retenu.

Partant du principe que le scénario initial trouvé par *forward search* ne correspond pas forcément au scénario maximisant la vraisemblance (propos illustré dans la Figure 2.3, à gauche), l’idée initiale a été de parcourir les sommets à la recherche de ce maximum. La première méthode employée était un recuit simulé. La quantité de paramètres tel que les critères d’arrêt ou la loi de la *température* ainsi que la lenteur de cette métaheuristique ont eu raison de la motivation initiale à utiliser cette méthode.

Puisqu’il s’agit de définir un sous-ensemble de scénarios, l’heuristique présentée ci-après a été choisie pour sa simplicité ; au détriment de son temps de calcul. Ce temps de calcul reste tout de même raisonnable puisque la méthode typiquement ne parcourt pas l’entier du réseau que représente l’hypercube. L’hypercube est un objet passablement étudié ; en particulier en théorie des graphes, des réseaux et, comme corollaire, en informatique. Il représente la structure qui a motivé l’élaboration du calcul parallèle. On trouve donc dans la littérature de nombreux algorithmes pour parcourir les sommets et transmettre de l’information au travers de ce réseau de manière efficace. Les méthodes dérivées de l’informatique ont été écartées pour leur complexité et également en raison du fait que généralement, ces méthodes parcourent tout l’hypercube et que cela n’est ici pas nécessaire. Une autre direction comme la définition d’un arbre couvrant sur l’hypercube n’a pas été retenue non plus, car l’abandon d’un noeud aurait signifié également l’abandon de tous les noeuds *enfants* qui lui succède. L’heuristique proposée est donc simple et ne présente pas de caractéristique d’optimalité en temps de parcours, mais elle a montré

qu'elle fonctionnait pour déterminer un voisinage du scénario initial fourni par la méthode *forward search*.

Détermination d'un voisinage de scénarios probables

Point 0 : Initialiser l'ensemble des scénarios retenus au singleton composé du scénario initial. On notera N_S la cardinalité de cet ensemble.

Point I : Considérer l'ensemble des $n \cdot N_S$ premiers sommets voisins des sommets déjà inclus dans l'ensemble des scénarios retenus.

Point II : Evaluer la fonction de vraisemblance associée à chacun de ces sommets.

Point III : Retenir uniquement les sommets tels que l'estimation de la vraisemblance associée est supérieure ou égale à l'estimation de la vraisemblance associée au sommet initial.

Point IV : Retourner au **Point I** et cycliser jusqu'à ce que la cardinalité N_S cesse d'augmenter.

Des illustrations de la détermination de ce sous-ensemble sont présentées dans la Figure 2.3 et dans la Figure 2.8. Dans le premier cas, 42 scénarios sur un total de 2^{36} ont été retenus tandis que dans le second cas, 20 scénarios, incluant celui fourni par la première étape, sur un total de 2^{16} ont été retenus. On constate que ces cardinalités sont très faibles en comparaison de la cardinalité totale qui rendent quasi impossible le calcul exact en considérant tous les scénarios. Ci-après, quelques propriétés de l'estimateur basé sur ce sous-ensemble de scénarios sont présentées.

2.2.4 Propriétés de l'estimateur

Ci-dessous sont présentés les propriétés basiques de l'estimateur proposé.

Proposition 2.6 L'estimateur utilisé fourni par (2.11) possède comme distribution un mélange de lois gaussiennes.

Proposition 2.7 L'estimateur utilisé fourni par (2.11) est sans biais. Ce résultat découle de sa construction.

$$\mathbb{E}(\hat{Z}_S(s_0)) = \sum_{b \in S} \tilde{\omega}_b \alpha_b^t \mathbb{E}(V) = \sum_{b \in S} \tilde{\omega}_b \alpha_b^t \mu \mathbf{1} = \sum_{b \in S} \tilde{\omega}_b \mu = \mu, \quad \forall \mu \in \mathbb{R}. \quad (2.15)$$

Pour juger de la qualité de l'estimateur, on calcule le carré moyen de l'erreur. Le problème de la cardinalité totale des scénarios possibles refait surface et seule une approximation du carré moyen de l'erreur peut être calculée.

Proposition 2.8 Le carré moyen de l'erreur de prédiction de l'estimateur (2.7) s'écrit

$$\begin{aligned}\sigma_{\mathcal{B}}^2(s_0) &= \mathbb{E}\left(\left(\widehat{Z}(s_0) - Z(s_0)\right)^2\right) \\ &= \text{Var}\left(\widehat{Z}(s_0)\right) - 2\text{Cov}\left(\widehat{Z}(s_0), Z(s_0)\right) + \text{Var}\left(Z(s_0)\right) \\ &= \sum_{b_i, b_j \in \mathcal{B}} \omega_{b_i} \omega_{b_j} \alpha_{b_i}^t \Omega_{ij} \alpha_{b_j} - 2\sqrt{\sigma^2}(1 - \varepsilon) \sum_{b_i \in \mathcal{B}} \omega_{b_i} \alpha_{b_i}^t \nu + \sigma^2,\end{aligned}$$

où les termes α_{b_i} s'écrivent comme dans (2.10), pour tout $i = 1, \dots, 2^n$, $\varepsilon = \mathbf{P}(B(s_j) = 1)$, $\forall j = 1, \dots, n$ et

$$\Omega_{ij} = (\mathbf{I}_n - A_{b_i})\Sigma(\mathbf{I}_n - A_{b_j}) + k^2\sigma^2 A_{b_i} A_{b_j}, \quad \forall i, j = 1, \dots, 2^n.$$

Par conséquent, une approximation du carré moyen de l'erreur de prédiction de l'estimateur (2.11) est donné par

$$\sigma_S^2(s_0) \approx \sum_{b_i, b_j \in S} \omega_{b_i} \omega_{b_j} \alpha_{b_i}^t \Omega_{ij} \alpha_{b_j} - 2\sqrt{\sigma^2}(1 - \varepsilon) \sum_{b_i \in S} \omega_{b_i} \alpha_{b_i}^t \nu + \sigma^2.$$

Preuve 2.9 Il s'agit de calculer

$$\begin{aligned}\sigma_{\mathcal{B}}^2(s_0) &= \mathbb{E}\left(\left(\widehat{Z}_{\mathcal{B}}(s_0) - Z(s_0)\right)^2\right) \\ &= \mathbb{E}\left(\left(\widehat{Z}_{\mathcal{B}}(s_0) - \mathbb{E}(\widehat{Z}_{\mathcal{B}}(s_0) \mid Z(s_0)) + \mathbb{E}(\widehat{Z}_{\mathcal{B}}(s_0) \mid Z(s_0)) - Z(s_0)\right)^2\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(\left(\widehat{Z}_{\mathcal{B}}(s_0) - \mathbb{E}(\widehat{Z}_{\mathcal{B}}(s_0) \mid Z(s_0)) + \mathbb{E}(\widehat{Z}_{\mathcal{B}}(s_0) \mid Z(s_0)) - Z(s_0)\right)^2 \middle| Z(s_0)\right)\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(\left(\widehat{Z}_{\mathcal{B}}(s_0) - \mathbb{E}(\widehat{Z}_{\mathcal{B}}(s_0) \mid Z(s_0))\right)^2 \middle| Z(s_0)\right)\right) \\ &\quad + 2 \cdot \mathbb{E}\left(\mathbb{E}\left(\left(\widehat{Z}_{\mathcal{B}}(s_0) - \mathbb{E}(\widehat{Z}_{\mathcal{B}}(s_0) \mid Z(s_0))\right)\left(\mathbb{E}(\widehat{Z}_{\mathcal{B}}(s_0) \mid Z(s_0)) - Z(s_0)\right) \middle| Z(s_0)\right)\right) \\ &\quad + \mathbb{E}\left(\mathbb{E}\left(\left(\mathbb{E}(\widehat{Z}_{\mathcal{B}}(s_0) \mid Z(s_0)) - Z(s_0)\right)^2 \middle| Z(s_0)\right)\right).\end{aligned}$$

Les trois termes de la dernière égalité sont évalués séparément ci-dessous. Le premier peut

se réécrire comme suit.

$$\begin{aligned} \mathbb{E} \left(\mathbb{E} \left(\left(\hat{Z}_{\mathcal{B}}(s_0) \mid Z(s_0) - \mathbb{E}(\hat{Z}_{\mathcal{B}}(s_0) \mid Z(s_0)) \right)^2 \right) \right) &= \mathbb{E} \left(\text{Var}(\hat{Z}_{\mathcal{B}}(s_0) \mid Z(s_0)) \right) \\ &= \text{Var}(\hat{Z}_{\mathcal{B}}(s_0)) - \text{Var} \left(\mathbb{E}(\hat{Z}_{\mathcal{B}}(s_0) \mid Z(s_0)) \right), \end{aligned}$$

en utilisant l'identité

$$\text{Var}(Q) = \text{Var}(\mathbb{E}(Q \mid R)) + \mathbb{E}(\text{Var}(Q \mid R)).$$

D'une part,

$$\begin{aligned} \text{Var}(\hat{Z}_{\mathcal{B}}(s_0)) &= \text{Var} \left(\sum_{b \in \mathcal{B}} \omega_b \hat{Z}_b(s_0) \right) = \sum_{b_i, b_j \in \mathcal{B}} \omega_{b_i} \omega_{b_j} \text{Cov}(\hat{Z}_{b_i}(s_0), \hat{Z}_{b_j}(s_0)) \\ &= \sum_{b_i, b_j \in \mathcal{B}} \omega_{b_i} \omega_{b_j} \alpha_{b_i}^t \text{Cov}((\mathbf{I}_n - A_{b_i})Z + A_{b_i}O, (\mathbf{I}_n - A_{b_j})Z + A_{b_j}O) \alpha_{b_j} \\ &= \sum_{b_i, b_j \in \mathcal{B}} \omega_{b_i} \omega_{b_j} \alpha_{b_i}^t ((\mathbf{I}_n - A_{b_i})\Sigma(\mathbf{I}_n - A_{b_j}) + k^2 \sigma^2 A_{b_i} A_{b_j}) \alpha_{b_j} \\ &= \sum_{b_i, b_j \in \mathcal{B}} \omega_{b_i} \omega_{b_j} \alpha_{b_i}^t \Omega_{ij} \alpha_{b_j}. \end{aligned}$$

D'autre part, pour la deuxième partie du premier terme, on a besoin de connaître la valeur de l'espérance conditionnelle $\mathbb{E}(\hat{Z}(s_0) \mid Z(s_0))$. On peut écrire

$$\begin{aligned} \mathbb{E}(\hat{Z}(s_0) \mid Z(s_0)) &= \sum_{b \in \mathcal{B}} \omega_b \alpha_b^t \mathbb{E}(X \mid Z(s_0)) \\ &= \sum_{b \in \mathcal{B}} \omega_b \alpha_b^t \mathbb{E}(\mathbb{E}(V \mid Z(s_0))) \\ &= \sum_{b \in \mathcal{B}} \omega_b \alpha_b^t \left(\sum_{b \in \mathcal{B}} p_b \mathbb{E}(V \mid Z(s_0)) \right), \end{aligned}$$

où $p_b = \mathbf{P}(B = b)$, $\forall b \in \mathcal{B}$. Puisque

$$\mathbb{E}(V \mid Z(s_0)) = \mu \mathbf{1} + (\mathbf{I}_n - A_b)\nu \cdot \frac{Z(s_0) - \mu}{\sqrt{\sigma^2}},$$

on peut écrire

$$\begin{aligned} \mathbb{E}(X \mid Z(s_0)) &= \sum_{b \in \mathcal{B}} p_b \left(\mu \mathbf{1} + (\mathbf{I}_n - A_b)\nu \cdot \frac{Z(s_0) - \mu}{\sqrt{\sigma^2}} \right) \\ &= \mu \mathbf{1} + \left(\sum_{b \in \mathcal{B}} p_b (\mathbf{I}_n - A_b) \right) \nu \cdot \frac{Z(s_0) - \mu}{\sqrt{\sigma^2}}. \end{aligned}$$

Finalement, en remarquant que $\sum_{b \in \mathcal{B}} p_b (\mathbf{I}_n - A_b) = \mathbf{I}_n - \sum_{b \in \mathcal{B}} p_b A_b = (1 - \varepsilon) \mathbf{I}_n$, on peut conclure que

$$\mathbb{E}(\hat{Z}(s_0) | Z(s_0)) = \sum_{b \in \mathcal{B}} \omega_b \alpha_b^t \left(\mu \mathbf{1} + (1 - \varepsilon) \nu \cdot \frac{Z(s_0) - \mu}{\sqrt{\sigma^2}} \right). \quad (2.16)$$

Ainsi, le premier des trois termes à calculer se réécrit comme

$$\begin{aligned} \mathbb{E} \left(\mathbb{E} \left(\left(\hat{Z}_{\mathcal{B}}(s_0) - \mathbb{E}(\hat{Z}_{\mathcal{B}}(s_0) | Z(s_0)) \right)^2 \right) \middle| Z(s_0) \right) &= \sum_{b_i, b_j \in \mathcal{B}} \omega_{b_i} \omega_{b_j} \alpha_{b_i}^t \Omega_{ij} \alpha_{b_j} \\ &\quad - (1 - \varepsilon)^2 \left(\sum_{b \in \mathcal{B}} \omega_b \alpha_b^t \nu \right)^2. \end{aligned}$$

Le second se développe comme suit.

$$\begin{aligned} 2 \cdot \mathbb{E} \left(\mathbb{E} \left(\left(\hat{Z}_{\mathcal{B}}(s_0) - \mathbb{E}(\hat{Z}_{\mathcal{B}}(s_0) | Z(s_0)) \right) \left(\mathbb{E}(\hat{Z}_{\mathcal{B}}(s_0) | Z(s_0)) - Z(s_0) \right) \right) \middle| Z(s_0) \right) \\ = 2 \cdot \mathbb{E} \left(\mathbb{E} \left(\left(\hat{Z}_{\mathcal{B}}(s_0) | Z(s_0) - \mathbb{E}(\hat{Z}_{\mathcal{B}}(s_0) | Z(s_0)) \right) \left(\mathbb{E}(\hat{Z}_{\mathcal{B}}(s_0) | Z(s_0)) - Z(s_0) \right) \right) \right) \\ = 2 \cdot \mathbb{E} \left(\left(\mathbb{E}(\hat{Z}_{\mathcal{B}}(s_0) | Z(s_0)) - Z(s_0) \right) \cdot \mathbb{E} \left(\hat{Z}_{\mathcal{B}}(s_0) | Z(s_0) - \mathbb{E}(\hat{Z}_{\mathcal{B}}(s_0) | Z(s_0)) \right) \right) = 0. \end{aligned}$$

En utilisant le résultat fourni par (2.16), le troisième et dernier terme s'écrit

$$\begin{aligned} \mathbb{E} \left(\mathbb{E} \left(\left(\mathbb{E}(\hat{Z}_{\mathcal{B}}(s_0) | Z(s_0)) - Z(s_0) \right)^2 \right) \middle| Z(s_0) \right) &= \mathbb{E} \left(\left(\mathbb{E}(\hat{Z}_{\mathcal{B}}(s_0) | Z(s_0)) - Z(s_0) \right)^2 \right) \\ &= \mathbb{E} \left(\left(\sum_{b \in \mathcal{B}} \omega_b \alpha_b^t \left(\mu \mathbf{1} + (1 - \varepsilon) \nu (Z(s_0) - \mu) / \sqrt{\sigma^2} \right) - Z(s_0) \right)^2 \right) \\ &= \mu^2 \left(\sum_{b \in \mathcal{B}} \omega_b \alpha_b^t \mathbf{1} - 1 \right)^2 + \sigma^2 - 2\sqrt{\sigma^2} (1 - \varepsilon) \sum_{b \in \mathcal{B}} \omega_b \alpha_b^t \nu \\ &\quad + (1 - \varepsilon)^2 \left(\sum_{b \in \mathcal{B}} \omega_b \alpha_b^t \nu \right)^2. \end{aligned}$$

En remarquant que $\sum_{b \in \mathcal{B}} \omega_b \alpha_b^t \mathbf{1} = 1$, le carré moyen de l'erreur de prédiction se réécrit comme

$$\sigma_{\mathcal{B}}^2 = \sum_{b_i, b_j \in \mathcal{B}} \omega_{b_i} \omega_{b_j} \alpha_{b_i}^t \Omega_{ij} \alpha_{b_j} - 2\sqrt{\sigma^2} (1 - \varepsilon) \sum_{b_i \in \mathcal{B}} \omega_{b_i} \alpha_{b_i}^t \nu + \sigma^2.$$

◆

Pour se convaincre que le terme croisé correspond bel et bien à la covariance entre $\widehat{Z}(s_0)$ et $Z(s_0)$, on peut développer le terme comme suit.

$$2\sqrt{\sigma^2}(1 - \varepsilon) \sum_{b \in \mathcal{B}} \omega_b \alpha_b^t \nu \propto \sum_{b \in \mathcal{B}} \omega_b \left(\nu^t (\mathbf{I}_n - A_b) \Omega_b^- \left(\nu - \frac{\mathbf{1}^t \Omega_b^- \nu}{\mathbf{1}^t \Omega_b^- \mathbf{1}} \cdot \mathbf{1} \right) + \frac{\mathbf{1}^t \Omega_b^- \nu}{\mathbf{1}^t \Omega_b^- \mathbf{1}} \right).$$

Dans cette dernière expression, le second terme peut être vu comme une moyenne pondérée des termes $\text{Cov}(Z(s_i), Z(s_0))$, $i = 1, \dots, n$, où la pondération est fonction des termes $\text{Cov}(Z(s_i), Z(s_j))$, $i, j = 1, \dots, n$, au travers de la matrice Ω_b^- . Le premier terme représente la norme L^2 au carré des écarts entre les termes $\text{Cov}(Z(s_i), Z(s_0))$, $i = 1, \dots, n$ et le centre que représente le second terme. Cette norme utilise également la matrice Ω_b^- pour pondération et seuls les termes non contaminés dans le scénario b participent au calcul de cet écart.

Proposition 2.10 Pour comparaison, la variance de prédiction associée au krigeage ordinaire est donnée par

$$\sigma_{\text{ok}}^2 = \lambda^t \left(\sum_{b \in \mathcal{B}} p_b \Omega_b \right) \lambda + \sigma^2 - 2\sqrt{\sigma^2}(1 - \varepsilon) \lambda^t \nu, \quad (2.17)$$

où $p_b = \mathbf{P}(B = b)$, $\forall b \in \mathcal{B}$ et λ^t est donné par (1.15).

Les questions qui se posent maintenant sont les suivantes :

1. Que vaut l'estimateur $\widehat{Z}_S(s_0)$ fourni par (2.11) comparé à $\widehat{Z}_B(s_0)$ fourni par (2.7) ?
2. La variance $\sigma_S^2(s_0)$ fournit-elle des valeurs sensées en comparaison de $\sigma_B^2(s_0)$?
3. Que vaut l'estimateur développé en comparaison de méthodes plus classiques telles que le krigeage robuste ou la médiane pondérée, toutes deux présentées dans le Chapitre 1 ?

Pour se rendre compte des performances de l'estimateur, des simulations ont été effectuées. Elles sont présentées dans le paragraphe suivant.

2.2.5 Simulations

Afin d'établir la qualité de l'estimateur présenté dans la section précédente, des simulations sont exécutées. Par simplicité, le domaine \mathcal{D} considéré est défini comme le carré unitaire $[0, 1] \times [0, 1]$. Sur ce domaine, on appose une grille régulière de $9 \times 9 = 81$ points et les prédictions ont toujours lieu au centre du domaine, au site dont les coordonnées sont $s_0 = (0.5, 0.5)^t$. La structure spatiale présente dans le processus d'intérêt provient d'un variogramme sphérique. Le choix du variogramme provient du fait que l'étendue finie de ce dernier fournit l'assurance d'un ajustement sans problème de convergence et permet,

par conséquent d'être inclu dans une boucle de simulation en toute confiance. Pour ces simulations, un nombre de $r = 1000$ répétitions a été fixé. Différentes situations correspondant à différentes structures spatiales sont considérées. Les paramètres qui varient au cours de ces simulations sont :

L'étendue du variogramme : $\theta_3 \in \{0.5, 0.7, 0.9\}$,

Le taux de contamination : $\epsilon \in \{0.00, 0.05, 0.10, 0.15\}$.

Les autres paramètres sont constants tout au long des simulations. Leurs valeurs sont :

La tendance : $\mu = 0$,

Le facteur de proportionnalité : $k^2 = 10$,

L'effet pépite du variogramme : $\theta_1 = 0.05$,

Le seuil du variogramme : $\theta_2 = 0.95$,

Le niveau du test de détection : $1 - \alpha = 0.9$.

L'estimateur développé et présenté dans la section précédente est comparé à plusieurs autres estimateurs. La dépendance spatiale n'est fournie à aucun d'entre eux. Par contre, elle est estimée une seule fois et l'estimation obtenue est utilisée par tous les estimateurs la nécessitant. Lors de l'ajustement du variogramme par la méthode des moindres carrés ordinaires, la famille paramétrique utilisée est fournie. Autrement dit, lors des simulations, c'est un variogramme sphérique qui est utilisé pour générer les données et c'est également un variogramme sphérique qui est ajusté. L'estimation se fait par la méthode proposée par Genton (1998a). Concernant les localisations des sites contaminés, elles sont aléatoires sur la grille comme supposé dans le modèle. Les différents estimateurs comparés dans ces simulations sont présentés ci-après.

FS : L'estimateur construit sur le scénario de contamination obtenu par l'automatisation de la méthode de Cerioli and Riani (1999). Cette automatisation a été présentée dans la Section 2.2.2.

SOK : L'estimateur présenté tout au long de ce chapitre et donné par (2.11).

RK : L'estimateur de krigeage robuste présenté dans la Section 1.3.1. Le paramètre de seuil pour l'édition des observations a été fixé à $c = 2$.

OK : L'estimateur de krigeage ordinaire donné par (1.16).

WAD : Cet acronyme vient de *weighted absolute deviation*. Cet estimateur correspond à une médiane pondérée des observations. Les poids utilisés sont ceux de l'estimateur de krigeage ordinaire, également utilisés par l'estimateur de krigeage robuste, une fois les observations éditées.

LMN : C'est une moyenne locale pondérée des 8 premiers sites voisins du site de prévision. Les poids utilisés sont proportionnels aux distances séparant les 8 sites du site de prévision.

LMD : C'est une médiane locale pondérée des 8 premiers sites voisins du site de prévision. Les poids utilisés sont proportionnels aux distances séparant les 8 sites du site de prévision.

L'estimateur fourni par (2.7) étant l'estimateur qui minimise le carré moyen de l'erreur pour ce modèle, on souhaite qu'une approximation de cet estimateur, fournie par (2.11) et dénommée dans cette section SOK, représente l'estimateur présentant les meilleures performances lors de ces simulations. La seconde remarque que l'on peut faire est que les performances de l'estimateur de krigeage robuste, RK, sont toutes relatives, car elles dépendent du paramètre de seuil c . En effet, étant donné que le processus étudié est entre autre stationnaire au premier ordre, l'erreur de prédiction de cet estimateur est en relation directe avec ce paramètre qui fixe la quantité d'observations éditées. C'est le seul estimateur parmi ceux présentés qui possède ce degré d'artificialité.

			Méthodes							
			FS	SOK	RK	OK	WAD	LMN	LMD	
Etendue	0.5	Contamination	0	0.496	0.491	0.501	<i>0.492</i>	0.575	0.507	0.527
			5	<i>0.498</i>	0.494	0.512	0.568	0.571	0.545	0.531
			10	<i>0.525</i>	0.521	0.549	0.627	0.583	0.589	0.534
			15	0.571	<i>0.548</i>	0.560	0.686	0.574	0.612	0.530
	0.7		0	0.415	0.410	0.427	<i>0.411</i>	0.531	0.430	0.446
			5	<i>0.414</i>	0.407	0.424	0.496	0.510	0.464	0.441
			10	0.453	0.439	<i>0.450</i>	0.561	0.522	0.521	0.453
			15	0.471	0.454	<i>0.464</i>	0.598	0.531	0.556	0.466
	0.9		0	0.379	0.372	0.384	<i>0.373</i>	0.473	0.383	0.389
			5	<i>0.377</i>	0.373	0.384	0.455	0.464	0.431	0.390
			10	0.420	0.406	<i>0.408</i>	0.530	0.497	0.484	0.409
			15	0.445	<i>0.425</i>	0.427	0.564	0.457	0.512	0.412

Tableau 2.2. Erreur absolue moyenne (MAE) obtenue par simulation. La valeur minimale parmi tous les estimateurs est en gras ; la seconde plus petite, en italique. Le processus de contamination est également gaussien avec un facteur de proportionnalité de la variance égal à $k^2 = 10$. Les localisations des sites contaminés sont aléatoires. Taille de l'échantillon : $r = 1000$. L'écart-type associé est d'environ 2.4%.

Hormis LMN et LMD, les estimateurs considérées utilisent tous l'ensemble des observations pour effectuer la prévision, même si, comme présenté dans la Figure 1.3, l'amplitude des poids associés aux observations décroît rapidement avec la distance séparant les

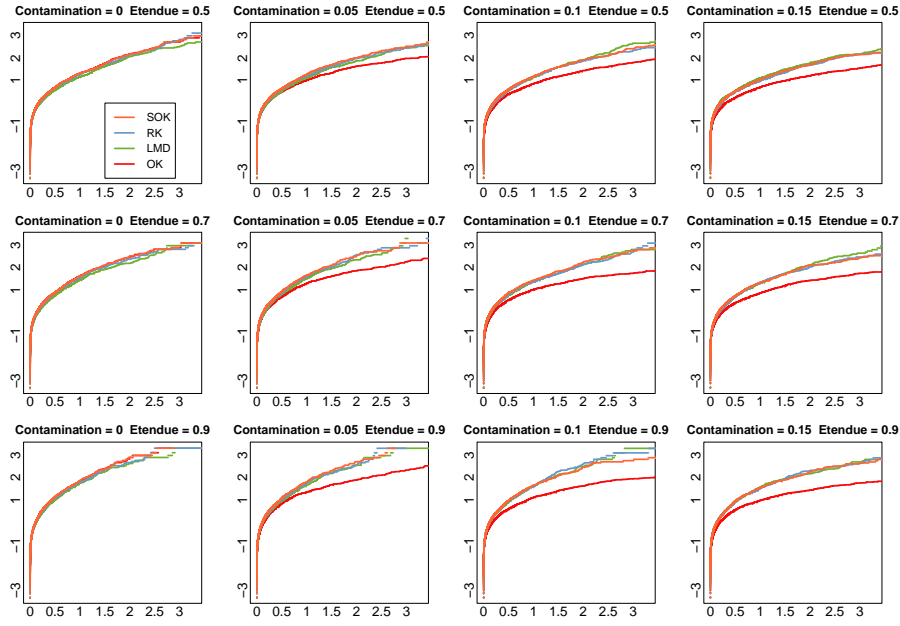


Figure 2.9. Comparaison des transformations probit des distributions empiriques de l'erreur absolue moyenne associée aux quatre estimateurs SOK (orange), RK (bleu), LMD (vert) et OK (rouge) dans les différentes situations de contamination normale envisagées.

sites. Par conséquent, à mesure que le paramètre d'étendue diminue, on s'attend à ce que les performances de l'estimateur de krigeage ordinaire et l'estimateur de moyenne locale soient de plus en plus similaires. Les deux statistiques considérées sont l'**erreur absolue moyenne**

$$\text{MAE} = \frac{1}{r} \sum_{i=1}^r |\hat{Z}(s_0) - Z(s_0)|$$

et la **racine du carré moyen de l'erreur**

$$\text{RMSE} = \sqrt{\frac{1}{r} \sum_{i=1}^r (\hat{Z}(s_0) - Z(s_0))^2}.$$

Les estimateurs des écarts-types associés à ces deux statistiques sont donnés par

$$\begin{cases} \hat{\sigma}_{\text{MAE}} = \frac{1}{r} \left(\sum_{i=1}^r (\hat{Z}(s_0) - Z(s_0))^2 - \frac{1}{r} \left(\sum_{i=1}^r |\hat{Z}(s_0) - Z(s_0)| \right)^2 \right)^{1/2} \\ \hat{\sigma}_{\text{RMSE}} = \left(\frac{1}{4r} \left(\frac{\sum_{i=1}^r (\hat{Z}(s_0) - Z(s_0))^4}{\sum_{i=1}^r (\hat{Z}(s_0) - Z(s_0))^2} - \frac{1}{r} \sum_{i=1}^r (\hat{Z}(s_0) - Z(s_0))^2 \right) \right)^{1/2} \end{cases}$$

			Méthodes							
			FS	SOK	RK	OK	WAD	LMN	LMD	
Etendue	0.5	Contamination	0	0.618	0.612	0.619	<i>0.612</i>	0.720	0.628	0.653
			5	<i>0.629</i>	0.625	0.644	0.740	0.713	0.685	0.664
			10	<i>0.658</i>	0.654	0.685	0.801	0.720	0.734	0.663
			15	0.737	<i>0.707</i>	0.708	0.897	0.729	0.781	0.675
	0.7		0	0.518	<i>0.511</i>	0.530	0.510	0.662	0.534	0.556
			5	<i>0.521</i>	0.513	0.534	0.669	0.641	0.596	0.551
			10	<i>0.570</i>	0.555	0.571	0.745	0.665	0.671	0.574
			15	0.608	0.581	<i>0.584</i>	0.796	0.668	0.720	0.590
	0.9		0	<i>0.471</i>	0.465	0.483	0.467	0.596	0.478	0.490
			5	<i>0.471</i>	0.469	0.484	0.596	0.586	0.546	0.490
			10	0.532	0.510	0.512	0.698	0.620	0.613	<i>0.510</i>
			15	0.583	<i>0.535</i>	0.538	0.762	0.584	0.671	0.528

Tableau 2.3. Racine de l'erreur carrée moyenne (RMSE) obtenue par simulation. La valeur minimale parmi tous les estimateurs est en gras ; la seconde plus petite, en italique. Le processus de contamination est également gaussien avec un facteur de proportionnalité de la variance égal à $k^2 = 10$. Les localisations des sites contaminés sont aléatoires. Taille de l'échantillon : $r = 1000$. L'écart-type associé est d'environ 2.4%.

Les résultats obtenus sont présentés dans les Tableaux 2.2 et 2.3 ainsi que dans la Figure 2.9. On constate les tendances suivantes :

- *A priori* l'estimateur SOK représente bel et bien l'estimateur linéaire sans biais qui minimise l'erreur carrée moyenne.
- Plus la surface du processus d'intérêt est lisse, autrement dit plus l'étendue du variogramme est grande, plus faibles sont le MAE et le RMSE de l'estimateur SOK.

Lorsque le taux de contamination est nul, l'estimateur SOK est similaire à l'estimateur OK, alors que lorsque le taux de contamination augmente, l'estimateur LMD est celui qui se rapproche le plus de l'estimateur SOK. Cet effet retranscrit dans le multidimensionnel ce qui avait déjà été observé dans l'exemple unidimensionnel d'introduction. En effet, si le taux de contamination est faible, l'estimateur aura tendance à faire confiance aux méthodes classiques et se comporter comme le krigeage ordinaire. A l'inverse, lorsque le taux de contamination est grand, l'estimateur proposé préfère s'éloigner des méthodes classiques et se rabattre sur une estimation inconditionnelle. Ceci se reflète par le faible écart séparant la médiane locale pondérée et l'estimateur dans cette situation.

Dans la Figure 2.9, les fonctions probit des distributions empiriques des quatre estimateurs SOK, RK, LMD et OK sont représentées pour chacune des situations envisagées.

Plus la contamination est forte, plus la différence entre l'estimateur proposé et le krigeage ordinaire se creuse, à l'avantage de l'estimateur. Si la contamination est faible, la distribution de l'estimateur est confondue avec celle du krigeage ordinaire. Encore une fois, il faut relativiser la performance du krigeage robuste dans le sens où celle-ci dépend directement du paramètre de seuil choisi.

Des simulations ont également été effectuées pour comparer les estimateurs de l'erreur de prédiction associés aux estimateurs comparés précédemment. Pour ce faire, des réalisations du modèle sous étude ont été simulées sur une grille unitaire régulière comme présentée dans la Figure 2.10. Le nombre de sites où le processus a été observé a spécialement été choisi comme faible ($n = 9$) afin de pouvoir calculer l'estimateur (2.11) dans le cas où ce n'est pas un sous-ensemble de scénarios qui est envisagé mais bien l'ensemble des scénarios possibles, ainsi que la variance associée. Les autres estimateurs considérés sont les estimateurs SOK, RK et OK. En observant, la Figure 2.10, on constate que le taux de contamination du processus est de $10/81 \approx 12\%$ et que parmi les sites observés, ce taux s'élève à $1/9 \approx 11\%$. L'étendue a été fixée à 0.9 (de sorte qu'il y ait une interaction entre chaque paire de points de la grille). Un total de $r = 1'000$ simulations ont été effectuées.

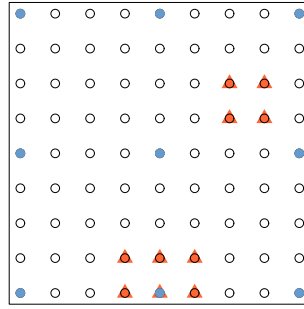


Figure 2.10. Grille pour les simulations concernant les estimateurs de l'erreur de prédiction. Sur cette grille régulière, les triangles (en orange) représentent les sites où le processus de contamination s'est réalisé ; en bleu, les sites où il y a eu observation.

Le premier point à préciser est que la variance de prédiction, elle aussi, dépend directement de l'estimation de la structure de dépendance. Pour s'en convaincre, il suffit de jeter un coup d'oeil à leur écriture pour les différents estimateurs ((2.8) et (1.3)). Ainsi, comme le précise Cressie (1993, p.127), le terme $\sigma^2(s_0)$ représente une erreur de prédiction carrée moyenne probabiliste et si l'on tient en compte le fait que le variogramme doit être estimé, alors l'inégalité

$$\sigma^2(s_0) \leq E\left(\left(\widehat{Z}(s_0) - Z(s_0)\right)^2\right) \quad (2.18)$$

est vérifiée qu'en probabilité lorsque les deux termes sont estimés. Idéalement, c'est le terme de droite qui devrait être utilisé pour quantifier l'erreur de prédiction en s_0 . Dans la pratique, c'est bien souvent une estimation du terme de gauche qui est utilisée. Ci-après, les deux termes de cette inégalité sont comparés pour les différents estimateurs.

Une comparaison graphique est proposée, puisque ces estimations varient pour chaque localisation et que pour ce simple exercice de simulation, il y a déjà 81 localisations à considérer. Dans la Figure 2.11, la variance de prédiction des estimateurs est présentée. Dans ce graphique, à gauche, l'estimateur basé sur l'ensemble des scénarios possibles ; au centre, l'estimateur SOK ; à droite le krigeage ordinaire. Dans la Figure 2.12, la somme moyenne des résidus au carré de l'estimateur basé sur l'ensemble des scénarios possibles, de l'estimateur SOK et du krigeage robuste (de gauche à droite).

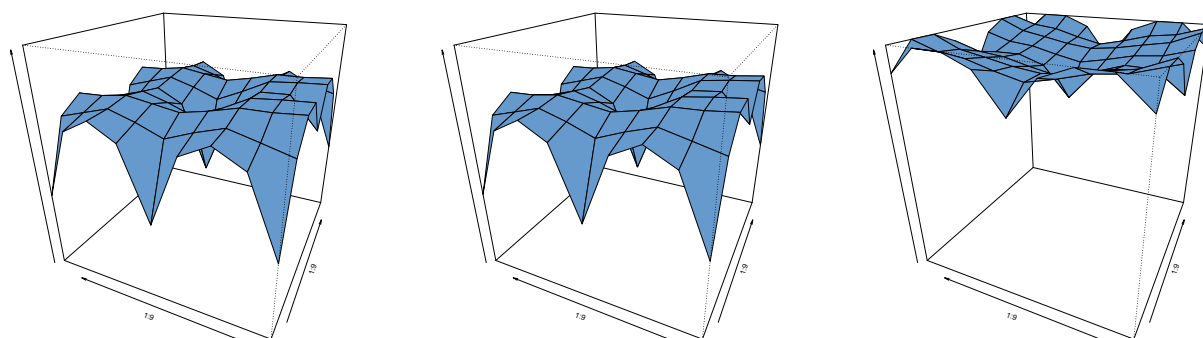


Figure 2.11. Comparaison des surfaces de variance de prédiction associées aux estimateurs. Dans l'ordre : à gauche, l'estimateur basé sur tous les scénarios possibles ; au centre, l'estimateur SOK ; à droite, l'estimateur OK.

Dans la Figure 2.11, la première observation est que la surface associée à l'estimateur SOK ressemble fortement à celle associée à l'estimateur basée sur l'ensemble des scénarios possibles. Toutes les deux sont nettement inférieures à celle associée au krigeage ordinaire. Il faut remarquer que ce sont des surfaces moyennes. En effet, contrairement à la variance de prédiction du krigeage ordinaire qui elle est constante au cours des simulations, la surface associée aux deux premiers estimateurs varie en fonction des observations. La différence entre les deux surfaces, celle associée à tous les scénarios et celle associée à la sélection, peut donc varier. On remarque également, que le krigeage ordinaire n'est plus un interpolateur exact dans ce cadre-ci. La pondération par le rapport de vraisemblance présente dans les deux autres estimateurs permet d'abaisser la variance de prédiction associée aux sites d'observation. Basé sur ces observations, on peut affirmer que d'une part l'estimateur SOK possède une variance de prédiction qui représente une bonne approximation de la variance de prédiction associée à l'estimateur basé sur l'ensemble des scénarios possibles et que d'autre part, sa supériorité face à des méthodes classiques est évidente. On peut encore remarquer que la formulation de la variance de prédiction associée au krigeage robuste n'a pas été dérivée pour cette situation, mais si on entreprend une analogie avec le modèle pour lequel le krigeage ordinaire est optimal, la surface associée au krigeage robuste est confondue avec celle associée au krigeage ordinaire.

Dans la Figure 2.12, les sommes moyennes des résidus au carré de prévisions sont présentées. On constate que les surfaces sont passablement similaires au niveau de leur forme. Par contre le pic présent, là où une contamination a effectivement été observée est beaucoup plus fort pour le krigeage robuste, tout à droite. Le pic associé au krigeage ordinaire était encore plus fort et n'est pas présenté pour une question d'échelle. On constate également que l'erreur moyenne ne faiblit pas si on se restreint à un sous-ensemble de scénarios, comme c'est le cas pour l'estimateur SOK, au centre, comparativement à celui de gauche qui est l'estimateur basé sur l'ensemble des scénarios.

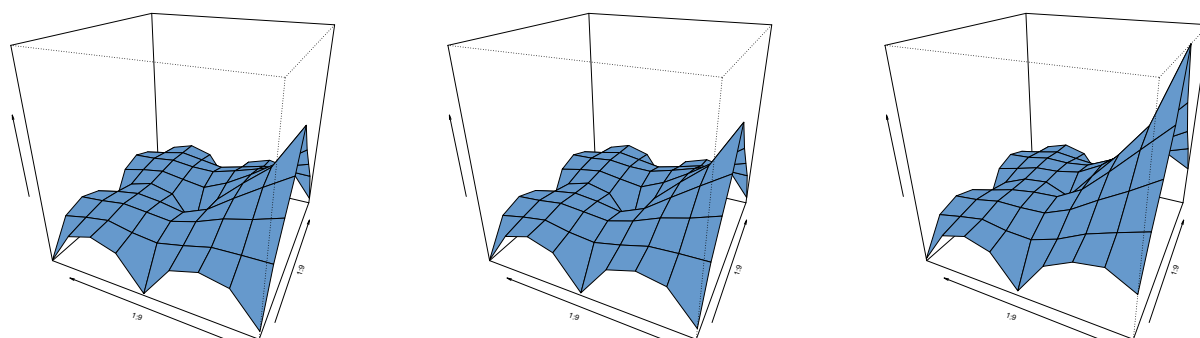


Figure 2.12. Comparaison des surfaces des sommes moyennes des résidus de prédiction au carré associées aux estimateurs. Dans l'ordre : à gauche, l'estimateur basé sur tous les scénarios possibles ; au centre, l'estimateur SOK ; à droite, l'estimateur RK.

Finalement, dans la Figure 2.13 et dans la Figure 2.14, une réalisation parmi les $r = 1'000$ a été considérée pour représenter la variance de prédiction et la somme des carrés des résidus de manière unidimensionnelle. Dans la première figure, on constate une fois encore que la variance de prédiction associée à SOK est confondue avec celle associée à l'estimateur basé sur l'ensemble des scénarios possibles, à l'exception de là où une contamination a eu lieu. Par symétrie, on peut voir que la courbe associée à SOK est la même, là où une contamination a été observée que là où aucune observation a été observée. A l'opposé, celle associée à tous les scénarios apporte du crédit également aux scénarios où le site contaminé est considéré comme propre. c'est ce qui fait que sa courbe se rapproche de celle du krigeage ordinaire. Dans la seconde figure, on constate que les résidus aux carrés associés à l'estimateur SOK sont nettement inférieurs à ceux associés au krigeage robuste ou à l'estimateur considérant tous les scénarios possibles.

Ce qu'il faut retenir de ces simulations est que l'estimateur proposé possède une forte capacité d'adaptation et ceci est retranscrit par de bonnes performances dans toutes les situations. Des simulations où le facteur de proportionnalité des variances a été fixé à $k^2 = 16$ ont également été effectuées. Les résultats sont sensiblement les mêmes que ceux présentés ici. Finalement, un histogramme est encore présenté dans la Figure 2.15. La valeur d'intérêt correspond au pourcentage des poids associés aux scénarios retenus par

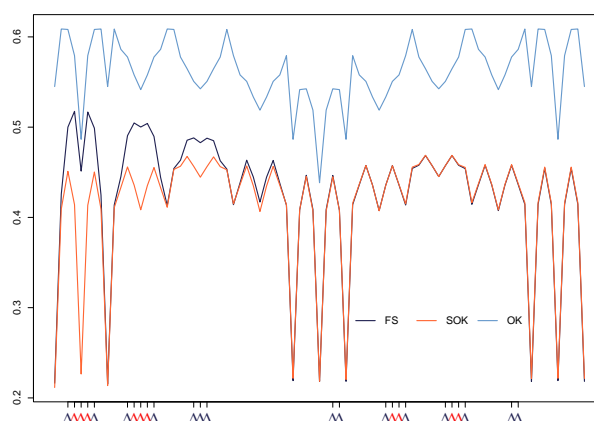


Figure 2.13. Représentation unidimensionnelle des variances de prédictions associées aux différents estimateurs. Les triangles rouges indiquent les sites posés comme contaminés dans le schéma de simulation ; les triangles noirs, leurs premiers voisins.

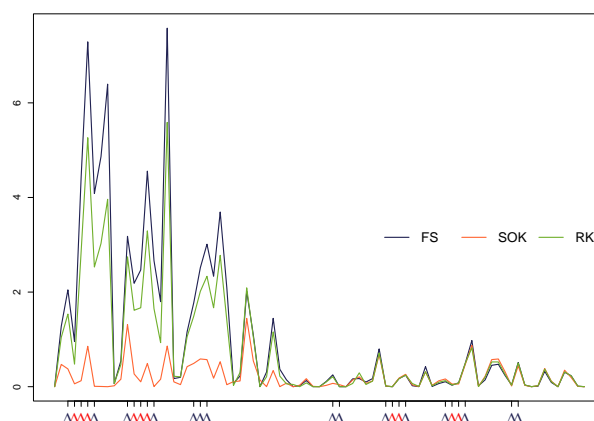


Figure 2.14. Représentation unidimensionnelle des résidus carrés associée à une prédiction des différents estimateurs. Le krigeage ordinaire n'est pas représenté pour une question d'échelle. Il est avantageusement remplacé par le krigeage robuste dont le comportement est similaire mais atténué.

l'estimateur SOK. On constate que la médiane se situe à environ 75%. Par conséquent, même si sa queue gauche se prolonge jusqu'à de très faibles valeurs, le rapport des poids se situe avec probabilité $p = 0.75$ au dessus de 68%. Dans la section suivante, une distribution à queues plus lourdes est envisagée pour le processus de contamination.

2.3 Contamination non gaussienne

Une des faiblesse du modèle présenté dans la Section 2.2 est que le processus d'intérêt $Z(\cdot)$ comme celui de contamination $O(\cdot)$ sont supposés gaussiens. Cette hypothèse a été faite dans le but de dériver facilement l'écriture de l'estimateur mais restreint tout de même son champ d'application. En particulier parce que la différence entre les deux dis-

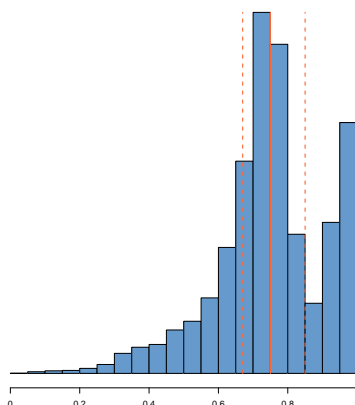


Figure 2.15. Histogramme du pourcentage associé aux poids des scénarios retenus par l'estimateur SOK.

tributions gaussiennes centrées dont les variances sont proportionnelles est minime (pour un facteur de proportionnalité raisonnable) à comparer au cas où une distribution gaussienne serait supposée pour le processus d'intérêt uniquement et qu'une distribution de Student serait considérée pour le processus de contamination. C'est cette motivation qui a poussé à vouloir relâcher cette hypothèse de normalité et à considérer des distributions elliptiques.

2.3.1 Distributions elliptiques

L'intérêt de ces lois réside dans le fait qu'elles permettent de s'éloigner un tant soit peu de l'hypothèse standard de normalité des données, tout en gardant des propriétés intéressantes qui les rendent relativement commodés à manipuler. En particulier, leur dénomination vient du fait qu'elles possèdent tout comme la loi normale des courbes d'isodensité elliptiques. Elles peuvent par contre exhiber des queues plus lourdes que la loi normale et par conséquent sont de bonnes candidates pour tester si la procédure mise en place est sensible ou non à l'hypothèse initiale sur la distribution des processus considérés. Une bonne introduction à ces distributions est proposée par Muirhead (1982, p.32) ou par Fang *et al.* (1990, p.26). Il existe plusieurs manières de définir une distribution elliptique. L'une consiste d'abord à définir une distribution sphérique.

Définition 2.11 Un vecteur aléatoire $X \in \mathbb{R}^n$ possède une **distribution sphérique** si pour toute transformation linéaire $H \in \mathbb{R}^{n \times n}$ orthogonale, X et $Y = HX$ possèdent la même distribution. En particulier, cela signifie que si f représente la densité de X , alors celle-ci peut s'écrire comme

$$f(x) = g(x^t x) = g(\|x\|),$$

pour une fonction g bien choisie.

Les distributions elliptiques représentent les transformations affines des distributions sphériques.

Définition 2.12 Soit Y un vecteur aléatoire distribué selon une loi sphérique avec pour densité $f(y) = g(y^t y)$. Soit $\mu \in \mathbb{R}^n$ et $A \in \mathbb{R}^{n \times n}$ régulière de sorte que $\det(A) \neq 0$ et que $\Sigma = A^t A$ soit strictement positive. Dans ce cas, le vecteur aléatoire $X = AY + \mu$ possède une **distribution elliptique** et sa densité s'écrit

$$h(x) = \frac{g((x - \mu)^t \Sigma^{-1} (x - \mu))}{\sqrt{\det(\Sigma)}}$$

et ses courbes d'isodensité sont des ellipsoïdes. Une distribution elliptique est complètement décrite par le paramètre de lieu μ , sa matrice de dispersion Σ et sa fonction caractéristique ϕ . On notera $X \sim E_\phi(\mu, \Sigma)$. Si le deuxième moment centré de X est défini, il est proportionnel à la matrice de dispersion Σ . Le facteur de proportionnalité, lui, vaut $-2\phi'(0)$.

Un exemple de distribution elliptique est donné par la distribution t multivariée avec p degrés de liberté. La densité s'écrit

$$t_p(x) = \frac{\Gamma(n+p)/2}{\Gamma(p/2)(p\pi)^{n/2}} \left(1 + \frac{x^t x}{p}\right)^{-(n+p)/2}, \quad x \in \mathbb{R}^n. \quad (2.19)$$

Ces distributions possèdent des propriétés qui les rendent intéressantes.

Proposition 2.13 Si $X \sim E_\phi(\mu, \Sigma)$ alors sa fonction caractéristique s'écrit

$$\phi(u) = e^{iu^t \mu} \psi(u^t \Sigma u)$$

avec ψ une fonction donnée.

Proposition 2.14 Soit $X \sim E_\phi(\mu, \Sigma)$ tel que

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

alors

$$E(X_1 | X_2) = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (X_2 - \mu_2), \quad (2.20)$$

$$\text{Var}(X_1 | X_2) = h(X_2) (\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}), \quad (2.21)$$

pour une fonction h donnée. De plus la distribution conditionnelle de $X_1 | X_2$ est encore

elliptique. Si la loi elliptique choisie est la loi normale, cette fonction est définie comme $h(X_2) \equiv 1, \quad \forall X_2$.

2.3.2 Le modèle

L'idée est maintenant de varier la forme de contamination grâce aux lois elliptiques tout en conservant la dépendance spatiale égale pour les deux processus considérés. La motivation à agir de la sorte vient du fait que la dépendance spatiale présente dans un processus sous étude dépend généralement moins du processus lui-même que de facteurs extérieurs tels que la topographie du domaine ou des agents atmosphériques par exemple. On suppose donc toujours être en présence d'un processus intrinséquement stationnaire se décomposant en deux parties ; le processus d'intérêt $Z(\cdot)$ et un processus de contamination substitutive $O(\cdot)$. Tous les deux possèdent des distributions elliptiques. En fait, pour $s_1, \dots, s_n \in \mathcal{D}$ on a que :

$$\begin{aligned} Z &= (Z(s_1), \dots, Z(s_n))^t \sim E_{\phi_Z}(\boldsymbol{\mu}, \Sigma), \\ O &= (O(s_1), \dots, O(s_n))^t \sim E_{\phi_O}(\boldsymbol{\mu}, \Sigma), \end{aligned}$$

où $\boldsymbol{\mu} = \mu \cdot \mathbf{1}$ et Σ est une matrice de dispersion strictement positive. Cette fois-ci, le processus de contamination possède également une dépendance spatiale. On suppose également que le processus d'intérêt possède un second moment fini, tandis que le processus de contamination, pas nécessairement. Finalement, le processus $B(\cdot)$ est supposé avoir toujours la même distribution que dans la section précédente, à savoir

$$B = (B(s_1), \dots, B(s_n))^t \sim \text{Bernoulli}(\varepsilon),$$

où $B(s_i) \perp B(s_j), \forall i \neq j$. Ces trois vecteurs sont indépendants entre eux, composante par composante, et sont toujours reliés entre eux par le modèle (2.4).

Afin de déterminer l'écriture de l'estimateur optimal (au sens L^2), on va comme dans la section précédente conditionner sur la connaissance du scénario de contamination. Posons $A_b = \text{diag}(b)$, où $b = (b(s_1), \dots, b(s_n))^t \in \{0, 1\}^n$. Puisqu'une variable aléatoire obtenue par transformation linéaire d'une autre variable aléatoire elliptique est encore elliptique, on obtient que

$$\begin{aligned} (\mathbf{I}_n - A_b)Z &\sim E_{\phi_Z}((\mathbf{I}_n - A_b)\boldsymbol{\mu}, (\mathbf{I}_n - A_b)\Sigma(\mathbf{I}_n - A_b)), \\ A_b O &\sim E_{\phi_O}(A_b\boldsymbol{\mu}, A_b\Sigma A_b). \end{aligned}$$

Puisque les processus $Z(\cdot)$ et $O(\cdot)$ sont indépendants, on peut conclure que le vecteur

aléatoire conditionné $V = X \mid (B = b) = (I_n - A_b)Z + A_b O$ possède la fonction caractéristique

$$\phi_V(t) = \phi_Z((I_n - A_b)t) \cdot \phi_O(A_b t) = e^{it^t \mu} \psi_Z(t^t \Sigma_Z t) \psi_O(t^t \Sigma_O t), \quad (2.22)$$

où $\Sigma_Z = (I_n - A_b)\Sigma(I_n - A_b)$ et $\Sigma_O = A_b \Sigma A_b$. De cette équation, on déduit la proposition présentée ci-après.

Proposition 2.15 Le processus V dont la fonction caractéristique est donnée par l'équation (2.22) n'est plus un processus elliptique.

Preuve 2.16 On peut appliquer une permutation aux sites d'observation de sorte que les sites non contaminés soient ordonnés l'un à la suite de l'autre et que les sites contaminés les suivent dans l'ordre. Si on appelle P_b la matrice associée à cette permutation, on obtient que

$$\begin{aligned} P_b(I_n - A_b)Z &\sim E_{\phi_Z}(P_b(I_n - A_b)\boldsymbol{\mu}, P_b(I_n - A_b)\Sigma(I_n - A_b)P_b^t), \\ P_b A_b O &\sim E_{\phi_O}(P_b A_b \boldsymbol{\mu}, P_b A_b \Sigma A_b P_b^t). \end{aligned}$$

La matrice $\tilde{\Sigma}_Z = P_b \Sigma_Z P_b^t$ est diagonale par bloc et de rang égal à $\text{rang}(\tilde{\Sigma}_Z) = \text{tr}(A_b)$. En fait, seul le premier bloc de cette matrice possède des composantes non nulles. De manière analogue, la matrice $\tilde{\Sigma}_O = P_b \Sigma_O P_b^t$ est diagonale par bloc et de rang égal à $\text{rang}(\tilde{\Sigma}_O) = n - \text{tr}(A_b)$. Dans cette matrice, seul le dernier bloc possède des composantes non nulles. Pour un scénario de contamination donné b et une matrice de permutation P_b associée, les matrices $\tilde{\Sigma}_Z$ et $\tilde{\Sigma}_O$ ne sont clairement pas identiques et

$$t^t \tilde{\Sigma}_Z t \neq t^t \tilde{\Sigma}_O t, \quad \forall t \in \mathbb{R}^n.$$

Par conséquent, la fonction caractéristique de $V(\cdot)$ n'est pas proportionnelle à une fonction d'une forme quadratique du type $t^t Q t$ pour une matrice Q donnée.

◆

Ci-dessous est énoncé un théorème sur lequel est basée la construction de l'estimateur proposé pour la prédiction de $Z(s_0)$.

Théorème 2.17 Soit $X(\cdot)$ un processus distribué selon un mélange de lois elliptiques comme présenté ci-dessus. Si on connaît le scénario de contamination $B = b$, l'estimateur optimal (au sens L^2) de $Z(s_0)$ basé uniquement sur $(I_n - A_b)X = (I_n - A_b)Z$ est donné par

$$\hat{Z}_b(s_0) = \left(\nu^t + \frac{1 - \nu^t M \mathbf{1}}{\mathbf{1}^t M \mathbf{1}} \cdot \mathbf{1}^t \right) M X, \quad (2.23)$$

où

$$M = (I_n - A_b)\Sigma_Z^-(I_n - A_b) \quad \text{et} \quad \Sigma_Z = (I_n - A_b)\Sigma(I_n - A_b).$$

Preuve 2.18 On cherche le vecteur de poids α_b tel que $\hat{Z}_b(s_0) = \alpha_b^t(I_n - A_b)Z$ qui vérifie la condition $\alpha_b^t(I_n - A_b)\mathbf{1} = 1$ et qui minimise le carré moyen de l'erreur. Ce dernier se réécrit comme

$$E((Z(s_0) - \hat{Z}_b(s_0))^2) = \sigma_Z^2 - 2\alpha_b^t(I_n - A_b)^t\nu + \alpha_b^t(I_n - A_b)\Sigma(I_n - A_b)\alpha_b.$$

En minimisant cette écriture par rapport aux poids α_b sous la contrainte que les poids somment à 1 on obtient le système d'équation :

$$\begin{cases} (I_n - A_b)\Sigma(I_n - A_b)\alpha_b = (I_n - A_b)(\nu + m\mathbf{1}), \\ \alpha_b^t(I_n - A_b)\mathbf{1} = 1, \end{cases}$$

où m représente le multiplicateur de Lagrange. En considérant l'inverse généralisé Σ_Z^- de Σ_Z , on obtient l'écriture de l'estimateur :

$$\hat{Z}_b(s_0) = \left(\nu^t + \frac{1 - \nu^t(I_n - A_b)\Sigma_Z^-(I_n - A_b)\mathbf{1}}{\mathbf{1}^t(I_n - A_b)\Sigma_Z^-(I_n - A_b)\mathbf{1}} \cdot \mathbf{1}^t \right) (I_n - A_b)\Sigma_Z^-(I_n - A_b)Z.$$

En posant $M = (I_n - A_b)\Sigma_Z^-(I_n - A_b)$, on obtient l'écriture proposée. L'espérance de l'estimateur vaut

$$\begin{aligned} E(\hat{Z}_b) &= E\left(\left(\nu^t + \frac{1 - \nu^t M \mathbf{1}}{\mathbf{1}^t M \mathbf{1}} \cdot \mathbf{1}^t \right) M X \right) \\ &= \mu \cdot \left(\nu^t + \frac{1 - \nu^t M \mathbf{1}}{\mathbf{1}^t M \mathbf{1}} \cdot \mathbf{1}^t \right) M \mathbf{1} \\ &= \mu \cdot \left(\nu^t M \mathbf{1} + \frac{\mathbf{1}^t M \mathbf{1}}{\mathbf{1}^t M \mathbf{1}} \mathbf{1}^t \mathbf{1} - \frac{\nu^t M \mathbf{1}}{\mathbf{1}^t M \mathbf{1}} \mathbf{1}^t \mathbf{1} \right) \\ &= \mu \end{aligned}$$

◆

Proposition 2.19 La variance de l'estimateur donné par (2.17) vaut

$$\text{Var}(\hat{Z}_b) = \left(\nu^t + \frac{1 - \nu^t M \mathbf{1}}{\mathbf{1}^t M \mathbf{1}} \cdot \mathbf{1}^t \right) M \left(\nu + \frac{1 - \nu^t M \mathbf{1}}{\mathbf{1}^t M \mathbf{1}} \cdot \mathbf{1} \right). \quad (2.24)$$

Basé sur le résultat du Théorème 2.17, on peut définir un estimateur raisonnable pour la prédiction de $Z(s_0)$ basé sur la connaissance de X . Même si l'idée est sensée,

prendre uniquement l'espérance de l'estimateur donné par (2.23) par rapport à B n'est pas intéressant puisque cet estimateur ne tiendrait pas compte dans sa pondération des valeurs observée. La définition ci-dessous présente l'estimateur considéré.

Définition 2.20 Dans le cas d'un mélange de lois elliptiques, un estimateur sensé pour la prédiction de $Z(s_0)$ sur la base de X est donné par

$$\widehat{Z}_B^e(s_0) = \sum_{b \in B} \omega_b \left(\nu^t + \frac{1 - \nu^t M \mathbf{1}}{\mathbf{1}^t M \mathbf{1}} \cdot \mathbf{1}^t \right) M X, \quad (2.25)$$

où

$$\omega_b = \mathbf{P}(B = b) \mathbf{P}(X = x \mid B = b) / \mathbf{P}(X = x), \quad (2.26)$$

et

$$M = (I_n - A_b) \Sigma_Z^- (I_n - A_b).$$

Par conséquent, un estimateur calculable en un temps raisonnable est donné par

$$\widehat{Z}_S^e(s_0) = \sum_{b \in S} \tilde{\omega}_b \left(\nu^t + \frac{1 - \nu^t M \mathbf{1}}{\mathbf{1}^t M \mathbf{1}} \cdot \mathbf{1}^t \right) M X, \quad (2.27)$$

où les poids $\tilde{\omega}_b$ sont proportionnels à la fonction de vraisemblance

$$\mathbf{P}(B = b) \mathbf{P}(X = x \mid B = b).$$

et sont tels que

$$1 \approx \sum_{b \in S} \tilde{\omega}_b \gg \sum_{b \notin S} \tilde{\omega}_b \approx 0.$$

Une fois l'ensemble S défini, les poids associés aux scénarios retenus sont normés pour assurer que l'estimateur obtenu soit toujours sans biais. En fait, cet estimateur correspond à une estimation de l'espérance de l'estimateur donné par (2.23) par rapport à la loi conditionnelle de B en sachant X .

Dans la section suivante, des simulations de cet estimateur comparé à d'autres estimateurs sont présentées.

2.3.3 Simulations

Le cadre choisi pour ces simulations est essentiellement le même que celui utilisé à la Section 2.2.5, où la loi de la contamination était normale. On considère une loi de Student de degré de liberté $p = 1$ donnée par (2.19) comme loi régissant le processus de contamination. Le processus d'intérêt reste lui gaussien. Les autres paramètres restent in-

				Méthodes						
				FS	SOK	RK	OK	WAD	LMN	LMD
Etendue	0.5	Contamination	0	0.494	0.482	0.489	<i>0.485</i>	0.562	0.492	0.511
			5	<i>0.487</i>	0.476	0.553	1.388	0.563	1.117	0.507
			10	0.645	0.610	<i>0.594</i>	0.983	0.621	0.744	0.559
			15	0.694	0.666	0.733	1.145	<i>0.598</i>	1.011	0.544
	0.7		0	0.412	0.409	0.428	<i>0.410</i>	0.522	0.423	0.438
			5	<i>0.428</i>	0.421	0.433	0.633	0.514	0.620	0.448
			10	0.548	<i>0.529</i>	0.773	1.831	0.534	1.516	0.463
			15	0.703	0.677	0.591	1.237	<i>0.544</i>	0.952	0.518
	0.9		0	0.371	0.364	0.379	<i>0.366</i>	0.472	0.379	0.394
			5	<i>0.387</i>	0.385	0.411	0.606	0.492	0.595	0.418
			10	0.479	<i>0.457</i>	0.487	0.891	0.503	0.880	0.441
			15	0.668	0.643	<i>0.488</i>	1.026	0.536	0.933	0.477

Tableau 2.4. Erreur absolue moyenne obtenue (MAE) par simulation. La valeur minimale parmi tous les estimateurs est en gras ; la seconde plus petite, en italique. Le processus de contamination est une loi de Student multivariée de degré $p = 1$. Les localisations des sites contaminés sont aléatoires. Taille de l'échantillon : $r = 1000$. L'écart-type associé est d'environ 2.8%.

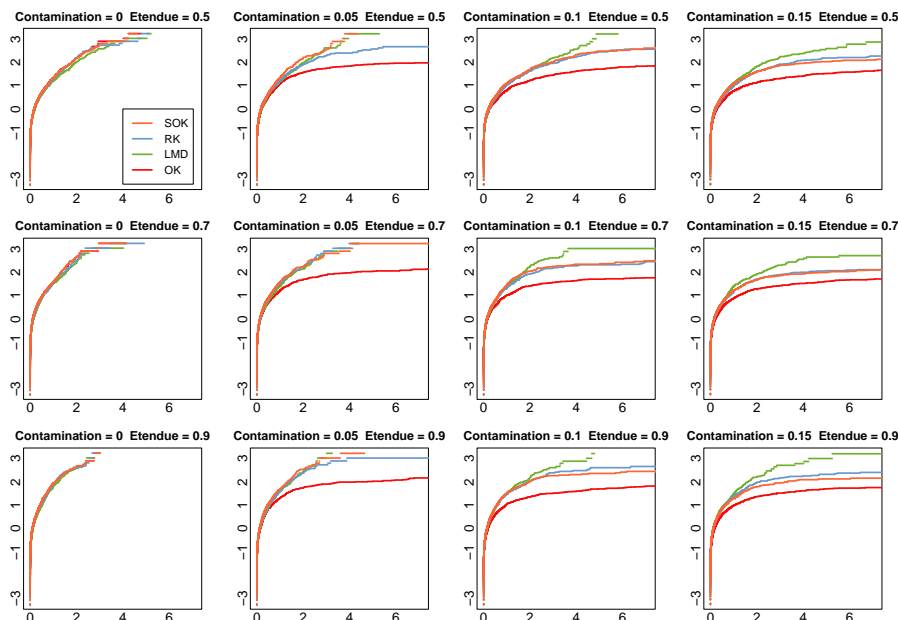


Figure 2.16. Comparaison des transformations probit des distributions empiriques de l'erreur absolue moyenne associée aux quatre estimateurs SOK (orange), RK (bleu), LMD (vert) et OK (rouge) dans les différentes situations de contamination de Student envisagées.

			FS	SOK	RK	OK	WAD	LMN	LMD	
Etendue	0.5	Contamination	0	0.611	0.595	0.597	<i>0.596</i>	0.705	0.614	0.634
			5	<i>0.614</i>	0.597	1.615	14.899	0.703	11.611	0.633
			10	1.835	1.745	0.807	4.760	<i>0.772</i>	1.672	0.702
			15	2.106	2.031	2.665	4.578	<i>0.751</i>	3.528	0.685
	0.7		0	0.518	0.515	0.534	<i>0.516</i>	0.653	0.527	0.548
			5	0.586	0.569	0.540	2.900	0.652	3.667	<i>0.561</i>
			10	1.918	1.832	6.061	24.511	<i>0.662</i>	17.446	0.581
			15	2.238	2.156	1.236	6.539	0.691	4.019	<i>1.172</i>
	0.9		0	0.470	0.464	0.476	<i>0.466</i>	0.596	0.475	0.490
			5	<i>0.488</i>	0.484	0.539	2.024	0.614	2.589	0.517
			10	0.793	0.696	1.029	3.697	<i>0.640</i>	5.323	0.561
			15	2.780	2.519	1.320	4.876	<i>1.840</i>	4.271	1.983

Tableau 2.5. Racine de l'erreur carrée moyenne (RMSE) par simulation. La valeur minimale parmi tous les estimateurs est en gras ; la plus petite, en italique. Le processus de contamination est une loi de Student multivariée de degré $p = 1$. Les localisations des sites contaminés sont aléatoires. Taille de l'échantillon : $r = 1000$. L'écart-type associé est d'environ 5.7%.

changés sauf le paramètre de seuil associé au krigeage robuste qui a été abaissé à $c = 1.5$. Les résultats sont présentés dans les Tableaux 2.4, 2.5 et dans la Figure 2.16. L'estimateur SOK possède toujours de bonnes propriétés pour autant que la contamination soit inférieure ou égale à 10%.

2.4 Application

La modélisation présentée dans les sections précédentes de ce chapitre ainsi que l'estimateur dérivé ont été utilisés dans le cadre d'un concours d'interpolation spatiale ouvert à toute la communauté scientifique. Intitulé *SIC04*, ce concours a pris place dans le courant de l'année 2004 (<http://www.ai-geostats.org/>). Ce concours représentait une chance unique de comparer d'une part la méthode proposée à d'autres méthodes compétitives et d'autre part de confronter la méthode à des données sinon réelles du moins potentiellement réelles. Le but de ce concours était de proposer un algorithme d'interpolation spatial automatisé pour le traitement et la surveillance de phénomène environnementaux. La principale difficulté provenait du fait que bien qu'automatisée, la méthode devait pouvoir traiter dans un temps raisonnable des données dites de routine ainsi que des données issues d'une situation de crise ou d'urgence. Les données étaient des mesures de radioactivité fournies par la plateforme européenne d'échange de données radiologiques

(<http://eurdeppub.jrc.it/>). Le concours consistait en deux jeux de données à traiter. Les mesures couvraient uniquement le territoire allemand et provenait par conséquent du système de surveillance allemand afin d'éviter des problèmes d'homogénéité dans les mesures. Le premier jeu de données ressemblait à une journée de mesure normale, semblable aux dix autres journées qui étaient préalablement fournies pour le calibrage des méthodes. Le second jeu de données était également issu d'une journée normale de mesure. Seulement, une petite région du domaine d'étude a été sélectionnée et un modèle de dispersion a été modélisé et simulé afin d'obtenir dans cette région des valeurs d'un ordre dix fois supérieur aux valeurs maximales présentes dans les autres jeux de données communiqués. Au total, $N = 1008$ stations ont été retenues. Les valeurs observées par $n = 200$ de ces stations, choisies aléatoirement, étaient communiquées aux participants, ainsi que leurs localisations. Le but consistait à prédire les valeurs observées par les autres 808 stations restantes. Les critères de qualité retenus étaient le ME (erreur moyenne), le MAE (erreur absolue moyenne), le RMSE (racine carrée de l'erreur carrée moyenne) et le coefficient de corrélation de Pearson. Les principaux résultats ainsi que des illustrations sont présentés ci-après. A noter que ce concours a donné lieu à deux publications, Fournier and Furrer (2005a) et Fournier and Furrer (2005b). Le concours a enregistré douze groupes de participants proposant au total 25 méthodes différentes provenant de domaines tels que la géostatistiques, les *splines*, les réseaux neuronaux et les *support vector machines*. Parmi les méthodes géostatistiques, les méthodes d'estimation utilisées étaient variées : krigeage ordinaire, krigeage universel et interpolation spatiale.

	Min.	Moyenne	Médiane	Max.	Ecart-type
Données complètes ($N = 1008$)	57.0	97.7	98.6	180.0	19.6
Données communiquées ($n = 200$)	58.2	96.2	97.6	153.0	17.6
Prévision par SOK ($N - n = 808$)	68.6	96.7	99.1	126.3	14.6

Tableau 2.6. Statistiques concernant le premier jeu de données

Pour notre part, on a supposé que le processus étudié pouvait s'écrire comme (2.1), à savoir comme la somme d'une tendance absorbant les variations à basse résolution et d'un processus stationnaire au second ordre s'occupant des variations à plus haute résolution. Pour estimer la structure de tendance, un modèle de régression locale a été ajusté (Cleveland, 1979) aux observations. Cette modélisation permet d'envisager des situations où les erreurs possèdent une distribution dont les queues sont plus étirées que celles de la loi normale. Une forme générale polynomiale a été choisie pour la tendance et les variables explicatrices étaient les coordonnées des sites d'observation. De plus, la surface devait être conditionnellement linéaire. C'est-à-dire qu'étant donné une coordonnée connue, la

surface de tendance ajustée était une fonction linéaire de la coordonnée restante. Il faut noter que la structure de corrélation du processus stationnaire au second ordre n'est pas entré en compte lors de l'ajustement de la surface de tendance. Etant donné que cette surface moyenne est sensée capturer les variations à haute échelle, il a été supposé qu'il n'y avait pas grand chose à gagner à l'incorporer. Une approche itérative aurait pourtant pu être envisagée. Pour l'ajustement aux données, la fonction `loess` du logiciel R a été utilisé et le paramètre de lissage a été fixé à 0.75. Une fois la tendance ajustée, il fallait ajuster le processus stationnaire au second ordre sur la base des résidus de la régression. Puisque les résidus des dix jeux de données d'entraînement présentaient une légère anisotropie dans la direction est-ouest, les coordonnées dans cette même direction ont été multipliées par un facteur de 1.1. La dépendance spatiale a été estimée puis ajustée de la même manière que dans les simulations présentées dans la Section 2.2.5, à la différence près que c'est une famille exponentielle qui a été retenue pour l'ajustement. La décision de ne pas considérer des variogrammes de Matern a été prise en raison de la difficulté à ajuster le paramètre de lissage associé difficilement compatible avec la contrainte d'automatisation de la méthode. Finalement, pour la détermination du scénario initial par la méthode présentée à la Section 2.2.2, le degré du quantile empirique considéré pour la procédure de seuil automatique apparaissant dans la détection des observations contaminées (voir (2.14)) a été fixé à $1 - \alpha = 0.9$.

	Min.	Moyenne	Médiane	Max.	Ecart-type
Données complètes ($N = 1008$)	57.0	106.1	98.9	1528.2	92.5
Données communiquées ($n = 200$)	58.2	109.0	97.9	1499.0	122.0
Prévision par SOK ($N - n = 808$)	68.7	97.5	99.4	127.2	15.2
Prévision par RK ($N - n = 808$)	69.0	99.3	99.2	209.7	19.9
Prévision par OK ($N - n = 808$)	59.4	104.9	98.7	476.4	45.3

Tableau 2.7. Statistiques concernant le second jeu de données.

A quelques exceptions près se trouvant parmi les méthodes basées sur les réseaux neuronaux, les performances des différentes méthodes étaient relativement similaires lorsqu'appliquées au premier jeu de données. Deux raisons à cela. Tout d'abord, le premier jeu de données communiqué ressemblait fortement aux dix communiqués précédemment pour l'étalonnage des méthodes. Ensuite, le premier jeu de données était un jeu de données *propre*. Dans les articles mentionnés plus haut deux autres méthodes étaient également comparées. Il s'agissait d'une méthode de krigeage basée sur les plus proches voisins (Cressie, 1993, p.158) et d'un interpolateur par moyenne simple, également appelée mesure globale de la tendance centrale (Cressie, 1993, p.370). Dans la première méthode, les huit plus proches voisins étaient retenus ; dans la seconde, seulement les quatre plus proches. Ici, la méthode proposée est comparée aux méthodes du krigeage ordinaire et du

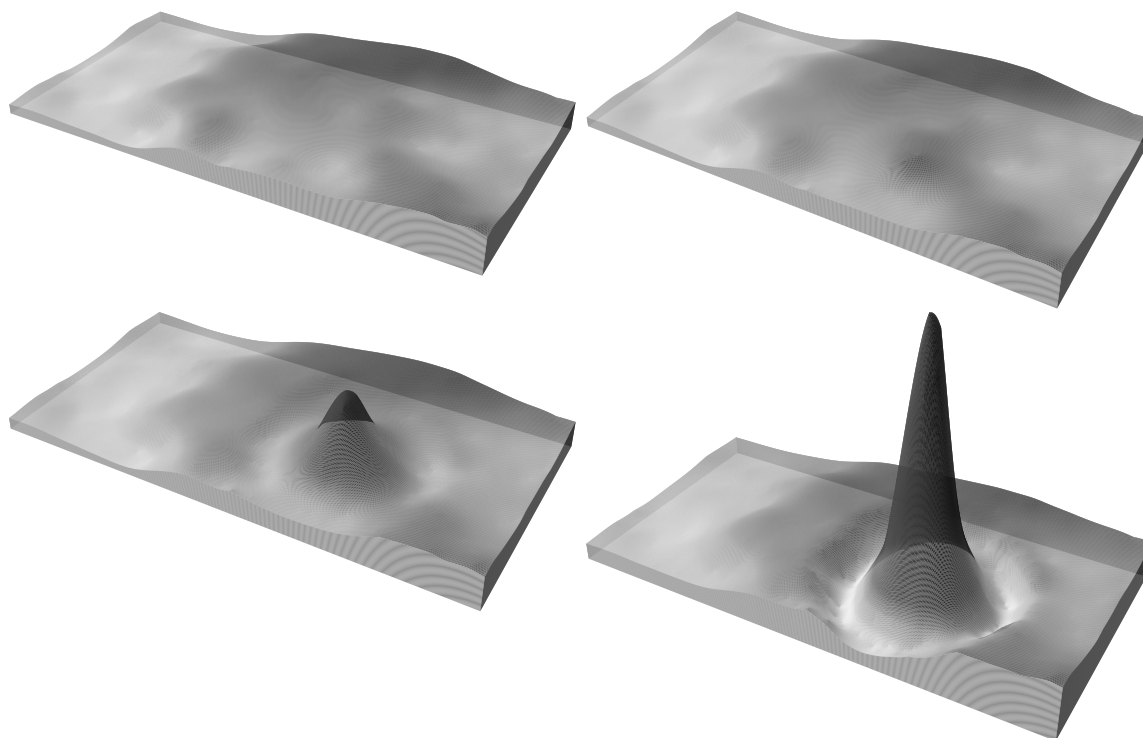


Figure 2.17. En haut, les surfaces obtenues en appliquant la modélisation à erreur substitutive au premier jeu de données (à gauche) et au second (à droite). En bas, les surfaces obtenues en appliquant le krigeage robuste (à gauche) et le krigeage ordinaire (à droite) au second jeu de données.

krigeage robuste tels que présentés au Chapitre 1.

	MAE	ME	Pearson's r	RMSE
SOK	9.11	-1.30	0.79	12.43
RK	9.12	-1.27	0.79	12.44
OK	9.12	-1.28	0.79	12.44
SOK	16.08	-7.92	0.31	80.69
RK	16.87	-6.14	0.44	77.19
OK	21.85	-0.49	0.48	73.15

Tableau 2.8. Comparaison de l'erreur pour le premier jeu de données (partie supérieure) et pour le second jeu de données (partie inférieure)

Dans le Tableau 2.6 sont présentées des statistiques concernant le premier jeu de données ainsi que la prévision obtenue grâce à la modélisation à erreurs substitutives. Les résultats sont quasi identiques pour les estimateurs de krigeage ordinaire et robuste que ceux obtenus à l'aide de la modélisation à erreur substitutive. Ceux concernant le second jeu de données sont donnés dans le Tableau 2.7. Quant aux valeurs prises par les différents critères de qualité, il sont présentés dans le Tableau 2.8.

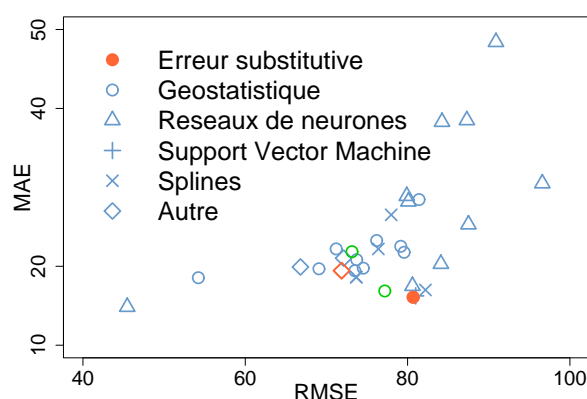


Figure 2.18. Comparaison de toutes les méthodes ayant participé à la compétition *SIC04* pour le second jeu de données. Sur l'axe des abscisses, le RMSE de chaque méthode ; sur l'axe des ordonnées, le MAE. En orange, la méthode SOK (rond) et l'interpolation par moyenne simple des quatre plus proche voisins (losange).

C'est avec le second jeu de données que les différentes méthodes ont pu être comparées. Lors de la compétition, la méthode présentée dans ce travail s'est distinguée en produisant le plus faible MAE et le plus faible RMSE pour le premier jeu de données. Pour le second jeu de données, une vue globale des résultats est présentée dans la Figure 2.18. La méthode à erreurs substitutives et celle d'interpolation par moyenne simple ont été mises en évidence. On remarque que le MAE associé à notre méthode est minimale parmi les méthodes de géostatistique et obtient la deuxième plus petite valeur parmi toutes les méthodes. Par contre, le RMSE n'est pas excellent. On pourrait arguer que ce critère, choisi par les organisateurs, n'est pas le meilleur pour juger des propriétés spécifiques de cet estimateur, mais son score reste acceptable. L'idée d'avoir proposé l'estimation par interpolation via une moyenne classique était également de montrer que si le but était de fournir une méthode automatique qui retransmette fidèlement le signal, il n'était pas nécessaire d'envisager quoi que ce soit de sophistiqué. La surface obtenue par cette méthode est représentée dans la Figure 2.19. On constate que cette méthode comparée à toutes les autres se comporte plutôt bien au niveau du MAE et du RMSE et c'est celle qui minimise l'erreur moyenne, estimateur du biais. Qui plus est, son temps de calcul se mesure en centième de seconde. La présentation de cet estimateur était donc également là pour emphaser le fait que mis à part les critères de qualité retenus par les organisateurs, il faut évaluer un estimateur en gardant à l'esprit la philosophie dans laquelle il a été mis en place.

L'idée de l'estimateur développé dans ce chapitre était de proposer un outil qui fournisse la meilleure estimation d'un phénomène lorsque son observation est "polluée" par un autre phénomène sans intérêt pour le scientifique. Traduit dans le contexte de la compétition, cela revient à dire que l'estimation obtenue représente la surface qu'on au-

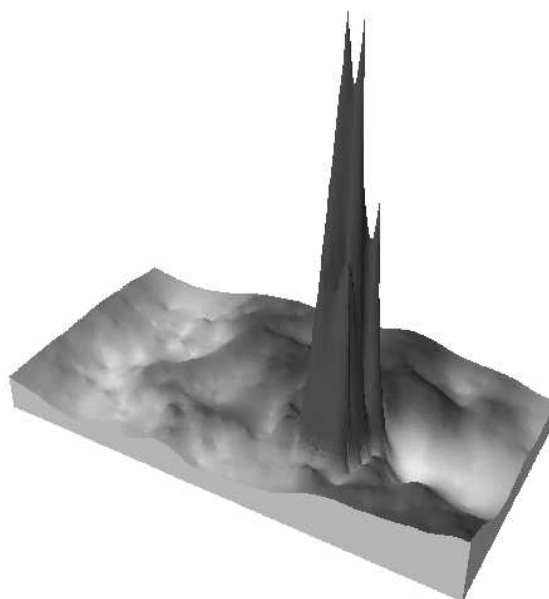


Figure 2.19. Représentation de la surface de prédiction obtenue en appliquant la méthode d'interpolation par moyenne simple sur les quatre plus proches voisins.

rait dû obtenir si il n'y avait pas eu d'accident ce jour-là. Certains pourraient arguer que le krigeage robuste propose déjà cette opportunité. Les Figures 2.17, 2.20 et 2.23 (page 73 pour cette dernière) illustrent ce propos et permettent de répondre à ces mêmes personnes. En effet, lorsque rien d'anormal se produit, il est difficile de distinguer entre les trois méthodes. Par contre, lorsque un accident survient, on constate tout d'abord que le krigeage ordinaire perd tout son intérêt. Encore une fois, cela dépend du but à atteindre. Mais si l'on souhaite retranscrire au mieux la surface, autant considérer l'interpolateur par moyenne simple. De la même manière, le krigeage robuste apparaît comme une concession, version atténuée du krigeage ordinaire. Une illustration de ce propos est fournie par la Figure 2.20. Tandis que l'estimateur proposé remplit son travail, dans le sens où les différences entre les surfaces associées aux deux jeux de données sont minimales. Au vu de ces résultats et de ces illustrations, l'estimateur proposé semble atteindre le but fixé.

Dans la prochaine section, une dernière modification de la modélisation est proposée. En effet, jusqu'alors il était supposé une absence de dépendance spatiale pour le processus de localisation de la contamination. Cette hypothèse est relâchée et les modifications entraînées par ce changement présentées.

2.5 Dépendance spatiale dans les scénarios

La dernière hypothèse contraignante présente dans le modèle est l'absence de dépendance spatiale dans $B(\cdot)$, le processus de localisation de la contamination. Dans la démarche

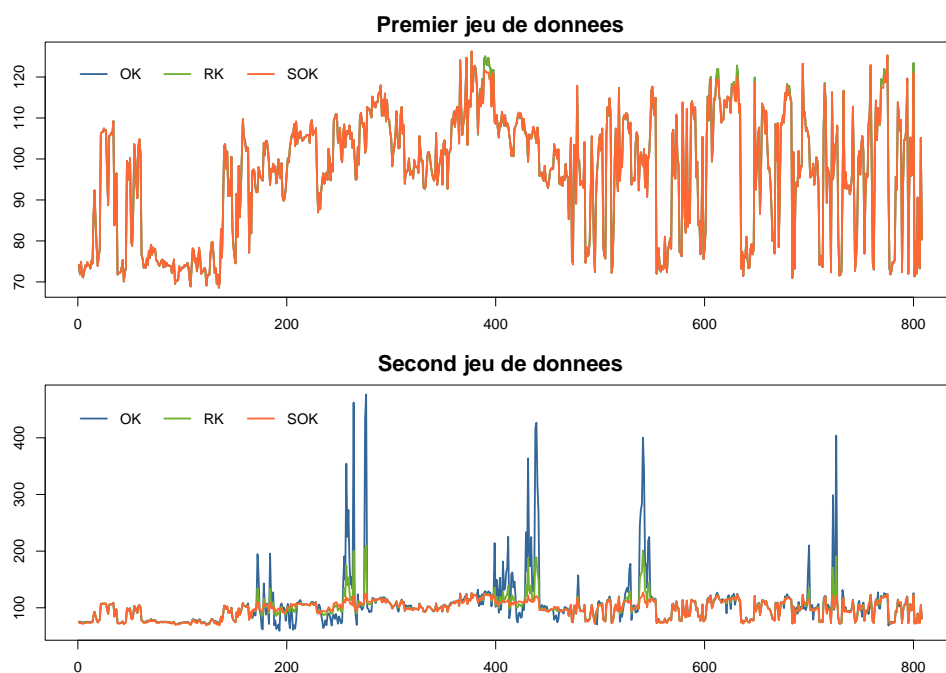


Figure 2.20. Représentation unidimensionnelle des prévisions obtenues par les trois méthodes envisagées : SOK (en orange), RK (en vert) et OK (en bleu).

proposée dans les sections précédentes, quelques changements ont été effectués pour pouvoir gérer la situation où la dépendance spatiale des localisations des sites contaminés n'est plus un bruit blanc.

Tout d'abord, dans la détermination du scénario de contamination le plus probable, selon la méthode *forward search* proposée par Cerioli and Riani (1999), ce n'est plus la suite de valeurs fournies par (2.12) qui est observée, mais celle donnée par

$$\sigma_m^{2(\max)} = \sigma_{[n](S_\star^{(m)})}^2, \quad m = p + 1, \dots, n - 1. \quad (2.28)$$

En effet, cette suite est sensée présenter un fort décroissement une fois que la première valeur contaminée est ordonnée, en raison du phénomène de masquage. Une fois de plus, pour déterminer un changement brutal dans une série, il apparaît intuitif de considérer la suite obtenue par différentiation de la suite de base :

$$\tilde{\sigma}_m^{2(\max)} = \sigma_m^{2(\max)} - \sigma_{m-1}^{2(\max)}, \quad m = p + 1, \dots, n. \quad (2.29)$$

Basé sur cette suite, les règles de détermination de la contamination d'un site sont données ci-dessous.

Règle I : Le site qui correspond à l'élément $\check{\sigma}_m^{2(\max)}$ est suspect si

$$\check{\sigma}_m^{2(\max)} > q\hat{F}_{(\check{\sigma}_{m-1}^{2(\max)}, \dots, \check{\sigma}_p^{2(\max)})}(1 - \alpha), \quad (2.30)$$

pour $0 < \alpha < 1$, choisi. Autrement dit, on compare chacun des termes de cette suite à un quantile de la loi empirique associée aux termes le précédant dans la suite.

Règle II : Le premier site suspect (dans l'ordre obtenu à la section précédente) et tous les sites suivants sont déclarés contaminés.

Le changement présenté ci-dessus est mineur relativement au second, puisqu'il s'agit d'une modification du modèle. En effet, le processus $B(\cdot)$ présente maintenant une dépendance spatiale. En fait, on suppose que ce processus est intrinsèquement stationnaire. Autrement dit, qu'il est stationnaire au premier ordre,

$$E(B(s)) = \epsilon, \quad s \in \mathcal{D},$$

et qu'il possède un variogramme donné par

$$2\gamma_B(s_1 - s_2) = \text{Var}(B(s_1) - B(s_2)), \quad s_1, s_2 \in \mathcal{D}.$$

Dès lors, deux problèmes liés l'un à l'autre apparaissent. Tout d'abord, la nécessité d'estimer la dépendance spatiale traduite par l'existence du variogramme et dans un deuxième temps, l'approximation de la vraisemblance associée à un jeu de données. En effet, les poids ω_b fournis par (2.8) sont basés sur la vraisemblance. Maintenant qu'une dépendance spatiale est supposée pour $B(\cdot)$,

$$\mathbf{P}(B = b) \neq \epsilon^{n_1} \cdot (1 - \epsilon)^{n - n_1},$$

où $n_1 = \#\{b(s_i) = 1 : i = 1, \dots, n\}$ représente le nombre de sites contaminés dans le scénario b et n représente toujours la taille de l'échantillon.

Même si les n variables aléatoires $B(s_i)$, $i = 1, \dots, n$ sont supposées identiquement distribuées, la loi conjointe

$$\mathbf{P}(B(s_1), \dots, B(s_n))$$

est uniquement définie si toutes les loi marginales du type

$$\mathbf{P}(B(s_{i_1}), \dots, B(s_{i_k})), \quad \forall 1 \leq i_1 < \dots < i_k \text{ et } k = 1, \dots, n \text{ le sont.}$$

Par conséquent, un calcul exact de la fonction de vraisemblance est rendu impossible,

même pour un nombre raisonnable n de variables aléatoires. Comme présenté dans Cox and Reid (2004), cette situation arrive fréquemment et dans d'autres domaines aussi variés que la génétique ou l'étude des valeurs extrêmes. Dans tous ces cas où la vraisemblance ne peut être calculée, elle est approchée en utilisant une pseudo-vraisemblance. La vraisemblance par paires est un cas particulier de cette approche. Des exemples d'application de ces techniques au domaine du spatial peuvent être trouvés dans Besag (1974), Heagerty and Lele (1998) et Nott and Rydén (1999). Une autre illustration de ces méthodes à des données binaires corrélées est trouvé dans Kuk and Nott (2000). L'idée principale est donc d'approcher la fonction de vraisemblance et par conséquent son estimateur du maximum de vraisemblance par une fonction composite approchant la fonction de vraisemblance et son maximum. Ici, c'est la fonction de vraisemblance par paires qui a été utilisée. Etant donné un vecteur $B = (B(s_1), \dots, B(s_n))$ on peut définir les contributions à la vraisemblance de premier et de second ordre :

$$l_1(\theta; B) = \sum_i \log f(B(s_i); \theta),$$

$$l_2(\theta; B) = \sum_{i>j} \log f(B(s_i), B(s_j); \theta) - aql_1(\theta; B),$$

où la valeur de la constante a est généralement choisie comme solution d'un problème d'optimisation, mais peut également être motivé par d'autres raisons. Tout d'abord, si la distribution marginale ne dépend pas du paramètre θ , il apparaît naturel de supposer $a = 0$. Ce choix peut également être motivé par soucis de diminuer le degré de complexité de la fonction. Si l'on choisit le paramètre $a = 0$, cela correspond à prendre en compte toutes les distributions bivariées possibles et est par conséquent également appelée fonction de vraisemblance par paires. L'estimateur du maximum de vraisemblance est obtenu en maximisant la fonction $l_2(\theta; B)$ par rapport à θ . Pour effectuer l'évaluation de la vraisemblance par paires, il est nécessaire de connaître la distribution conjointe de tous les couples de variables aléatoires $(B(s_i), B(s_j))$, $1 \leq i, j \leq n$. Le Tableau 2.1 décrit le lien entre cette distribution bivariée et le variogramme du processus $B(\cdot)$.

Etant donné que le processus spatial $B(\cdot)$ est binaire, l'estimateur classique donné par (1.9) est appliqué à la réalisation binaire obtenue après application de la méthode de classification présentée en début de section aux données originales. Une fois la dépendance estimée, le variogramme est ajusté en optimisant le critère des moindres carrés ordinaires sur une famille de variogrammes valides, en l'occurrence la famille sphérique. Le variogramme ajusté, les probabilités conjointes du type

$$\mathbf{P}(B(s_i) = b_i, B(s_j) = b_j), \quad b_i, b_j \in \{0, 1\}, \quad 1 \leq i, j \leq n,$$

sont connues et la fonction de vraisemblance par paires associée au jeu de données peut

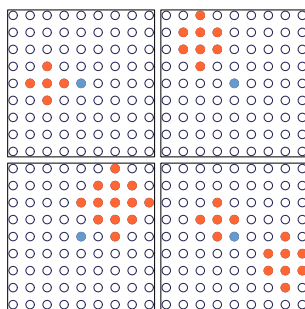


Figure 2.21. Représentation des quatre configurations de contamination envisagées. Le site en bleu représente la localisation de la prévision. Les sites en orange sont les sites contaminés. Leur localisation peut bouger dans la mesure où tous les sites contaminés sont toujours sur la grille et qu'ils n'empiètent pas sur le site de prévision. Les troisièmes et quatrièmes types de contamination présentent le même nombre de sites contaminés ; une fois en un seul groupe, l'autre fois en deux.

être évaluée. De la sorte, le voisinage du scénario de base peut également être déterminé et les estimateurs (2.11) ou (2.27) calculés. Des simulations ont été effectuées pour attester de la qualité de l'estimateur (2.27). Elles sont présentées dans la section suivante.

2.5.1 Simulations

Pour ces simulations, la distribution du processus de contamination a été choisie comme loi de Student avec un degré de liberté égal à 1. D'autre part, pour simuler des localisations de contaminations groupées, les échantillons ont été simulés selon l'un des quatre cas de figure présentés dans la Figure 2.21. La localisation du centre du groupe de sites contaminés était toujours aléatoire, mais une fois ce dernier décidé, ce sont les sites voisins selon l'un des quatres schémas illustrés qui sont contaminés. Les résultats sont présentés dans les Tableaux 2.9, 2.10 et dans la Figure 2.22. On constate que lorsque le processus de contamination se réalise de manière groupée, l'avantage de la méthode est évidente. Les seules méthodes concurrentes étant les médianes pondérées sur l'ensemble de la grille ou dans le voisinage du site de prédiction. Le krigeage robuste perd de son efficacité et ce, même si la valeur du seuil a été abaissée à $c = 1.5$.

			Méthodes							
			FS	SOK	RK	OK	WAD	LMN	LMD	
Etendue	0.5	Contamination	1	0.442	0.428	0.572	0.702	0.503	0.518	<i>0.437</i>
			2	0.445	0.407	0.608	0.858	0.482	0.756	<i>0.435</i>
			3	<i>0.667</i>	0.554	1.180	1.735	0.671	1.504	0.669
			4	0.490	0.477	0.670	0.918	0.498	0.758	<i>0.489</i>
	0.7		1	0.609	0.448	0.525	0.944	0.518	0.715	<i>0.479</i>
			2	0.526	0.488	0.720	0.851	0.574	0.660	<i>0.496</i>
			3	0.654	0.606	1.104	1.657	<i>0.636</i>	1.404	0.666
			4	0.531	0.507	0.746	1.209	0.538	0.897	<i>0.522</i>
	0.9		1	0.540	0.514	0.913	1.017	0.558	0.828	<i>0.515</i>
			2	0.548	0.518	0.883	1.799	0.569	0.960	<i>0.535</i>
			3	0.695	0.656	1.042	1.507	<i>0.659</i>	1.265	0.676
			4	0.664	<i>0.591</i>	1.157	2.031	0.593	1.312	0.558

Tableau 2.9. Erreur absolue moyenne obtenue (MAE) par simulation. La valeur minimale parmi tous les estimateurs est en gras ; la plus petite, en italique. Le processus de contamination est une loi de Student multivariée de degré $p = 1$. Les localisations des sites contaminés possèdent une dépendance spatiale. Taille de l'échantillon : $r = 1000$. L'écart-type associé est d'environ 5.3%.

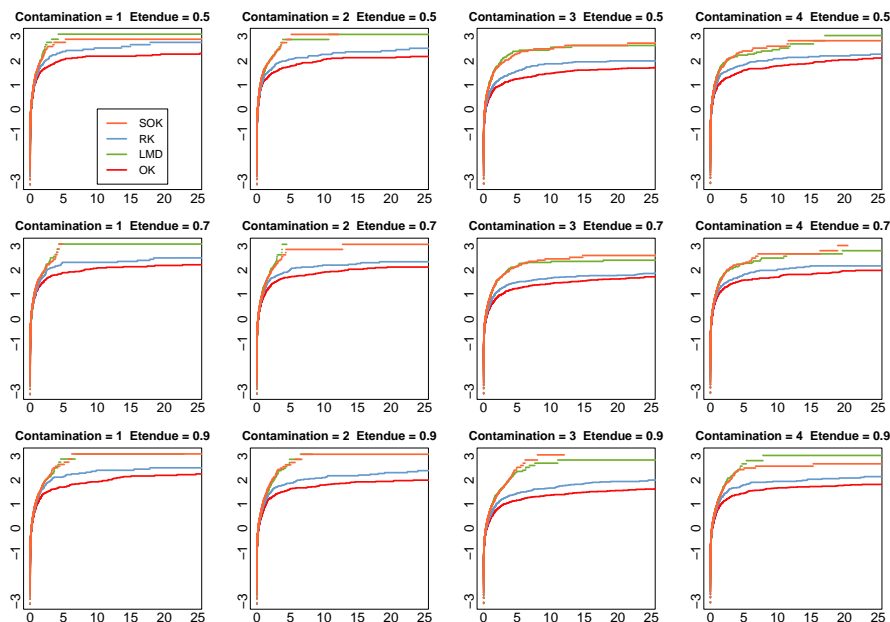


Figure 2.22. Comparaison des transformations probit des distributions empiriques de l'erreur absolue moyenne associée aux quatre estimateurs SOK (orange), RK (bleu), LMD (vert) et OK (rouge) dans les différentes situations de contamination de Student envisagées.

			Méthodes							
			FS	SOK	RK	OK	WAD	LMN	LMD	
Etendue	0.5	Contamination	1	0.780	<i>0.635</i>	2.878	3.762	0.737	1.061	0.535
			2	0.758	0.545	2.131	4.459	0.623	3.329	<i>0.607</i>
			3	1.777	<i>0.971</i>	4.646	8.421	0.705	7.044	3.628
			4	0.707	<i>0.684</i>	1.520	2.910	0.646	1.778	0.771
	0.7		1	3.198	0.576	0.872	8.464	<i>0.657</i>	4.178	0.666
			2	0.988	0.626	2.947	4.470	0.728	1.150	<i>0.627</i>
			3	1.047	<i>0.897</i>	4.147	8.517	0.798	5.497	1.308
			4	0.775	<i>0.692</i>	2.044	5.720	0.673	3.006	0.785
	0.9		1	0.833	<i>0.750</i>	7.250	6.127	0.775	4.100	0.677
			2	0.843	<i>0.729</i>	6.287	21.652	0.740	4.827	0.686
			3	1.025	0.828	2.537	4.564	<i>0.836</i>	3.399	1.129
			4	1.687	1.308	6.228	17.959	<i>0.732</i>	9.701	0.696

Tableau 2.10. Racine de l'erreur carrée moyenne obtenue (RMSE) par simulation. La valeur minimale parmi tous les estimateurs est en gras ; la plus petite, en italique. Le processus de contamination est une loi de Student multivariée de degré $p = 1$. Les localisations des sites contaminés possèdent une dépendance spatiale. Taille de l'échantillon : $r = 1000$. L'écart-type associé est d'environ 8%.

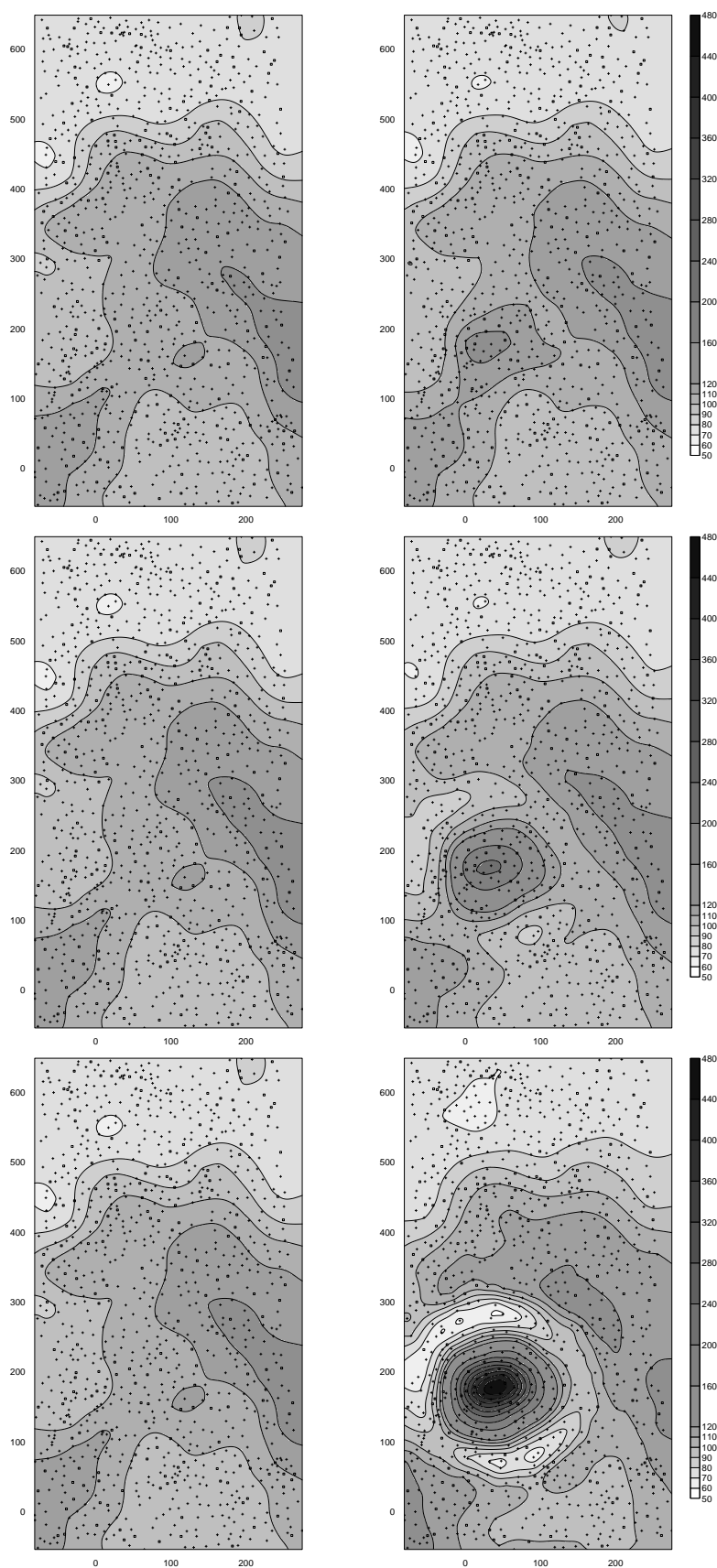


Figure 2.23. Comparaison des différentes surfaces obtenues pour le premier jeu de données (à gauche) et le second (à droite). En haut, la modélisation à erreur substitutive ; au centre, le krigeage robuste et au fond, le krigeage ordinaire.

Drap stable

En géostatistique, les données exhibent fréquemment simultanément des propriétés d'hétérogénéité et de similarité à différentes échelles d'étude. Par exemple, Painter (1995) montre comment des relevés de puits dans des champs pétroliers présentent le même comportement à différentes échelles, limitées vers le haut par la longueur finie des séquences de mesures et vers le bas par la définition des appareils de mesure. Il explique également que ni les propriétés des roches étudiées, ni les accroissements de ces mêmes propriétés ne présentent des comportements gaussiens, autrement dit homogènes. D'où l'idée d'une part de considérer des distributions qui permettent d'expliquer cette hétérogénéité et des modélisations qui capturent la propriété de similarité à différentes résolutions. C'est pourquoi, dans ce chapitre nous allons nous appliquer à développer une modélisation basée sur un drap stable. Dans un premier temps, les distributions infiniment divisibles et, en particulier, les distributions stables seront présentées. Ensuite, on présentera les processus stochastiques de Lévy et finalement, un modèle stochastique spatial sera mis en place et étudié.

3.1 Distributions infiniment divisibles

On considère un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$, où \mathcal{F} est une σ -algèbre construite sur les sous-ensembles de Ω et \mathbf{P} est une mesure de probabilité sur (Ω, \mathcal{F}) . Toute variable aléatoire Z unidimensionnelle est vue comme une forme \mathbf{P} -mesurable sur Ω . La loi de X

est notée μ_X , tel que pour tout $A \in \Omega$, $\mu_X(A) = \mathbf{P}(X \in A)$.

Définition 3.1 Les n variables aléatoires X_1, \dots, X_n sont dites indépendantes si pour tout choix $i_1, \dots, i_n \in \mathbb{N}$, $A_{i_1}, \dots, A_{i_n} \in \mathcal{B}(\mathbb{R})$ on a que $\mu_{X_1 \dots X_n}(A_1, \dots, A_n) = \prod_{j=1}^n \mu_{X_{i_j}}(A_{i_j})$, où $\mathcal{B}(\mathbb{R})$ représente l'ensemble des boréliens de \mathbb{R} .

La propriété suivante règle le cas de la somme de deux phénomènes aléatoire indépendants.

Propriété 3.2 La loi de la somme de deux variables aléatoires indépendantes est obtenue par convolution de leurs lois respectives.

Si on considère la suite $(S_n, n \in \mathbb{N})$ où

$$S_n = X_1^{(n)} + \dots + X_n^{(n)} \quad (3.1)$$

et que l'on suppose que

$$S_n \xrightarrow{D} S,$$

on peut se demander quelles sont les lois possibles pour S . Pour ce faire, on définit encore la caractéristique suivante.

Définition 3.3 Une variable aléatoire X possède une distribution infiniment divisible si pour tout $n \in \mathbb{N}$, il existe n variables aléatoires indépendantes et identiquement distribuées $X_1^{(n)}, \dots, X_n^{(n)}$ telles que

$$X \stackrel{D}{=} X_1^{(n)} + \dots + X_n^{(n)}. \quad (3.2)$$

On est maintenant à même de répondre à la question concernant la distribution de la limite de la suite $(S_n, n \in \mathbb{N})$ définie ci-dessus.

Proposition 3.4 Une variable aléatoire S représente la limite en distribution de sommes donnée par (3.1) si et seulement si sa distribution est infiniment divisible.

Une autre manière de définir la propriété de divisibilité infinie d'une distribution s'écrit en terme de la fonction caractéristique, transformée de Fourier de la densité. On peut rappeler que si les moments d'une variable aléatoire ne la définissent pas uniquement, la fonction caractéristique, au même titre que la distribution, la spécifie totalement. En fait, il existe une relation du type *un-un* entre les distributions et les fonctions caractéristiques.

Définition 3.5 Une fonction caractéristique ϕ est infiniment divisible si et seulement si pour tout $n \in \mathbb{N}$, il existe une fonction caractéristique ϕ_n telle que

$$\phi(u) = (\phi_n(u))^n, \quad \forall u \in \mathbb{R}.$$

Exemple 3.1.1

1. La fonction caractéristique d'une variable aléatoire $\mathcal{N}(\mu, \sigma^2)$ est donnée par

$$\phi(u) = \int_{x \in \mathbb{R}} e^{iux} (2\pi\sigma^2)^{-1/2} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} dx = e^{i\mu u - \frac{1}{2}\sigma^2 u^2}.$$

Par conséquent, la distribution normale est infiniment divisible, puisque pour tout $n \in \mathbb{N}$, elle peut être vue comme la distribution de la somme de n variables aléatoires indépendantes et identiquement distribuées $\mathcal{N}(\mu/n, \sigma^2/n)$.

2. Considérons maintenant N une variable aléatoire Poisson(λ). On pose

$$Z = X_1 + \dots + X_N,$$

où X_1, X_2, \dots sont des variables aléatoires indépendantes et identiquement distribuées. La variable aléatoire Z est dite Poisson composée et sa fonction caractéristique est donnée par

$$\phi_Z(u) = e^{\lambda(\phi_X(u)-1)},$$

où ϕ_X représente la fonction caractéristique des variables aléatoires X_n , $n \in \mathbb{N}$. La distribution Poisson composée est par conséquent infiniment divisible, puisque pour tout $n \in \mathbb{N}$, elle peut être vue comme la distribution de la somme de n variables aléatoires indépendantes et identiquement distribuées Poisson composée où les nombres de termes dans chaque somme N_1, \dots, N_n sont Poisson(λ/n).

En fait, on peut montrer que la classe des distributions infiniment divisibles coïncide avec la classe des distributions limites de distributions Poisson composées. La classe des distributions infiniment divisibles ne contient pas que des distributions Poisson composées (la loi normale est un contre-exemple), mais chacune d'entre elle représente la distribution limite d'une suite de distributions Poisson composées.

La fonction caractéristique est dans ce cadre un puissant outil, car à une distribution infiniment divisible ne peut correspondre qu'un certain type de fonction caractéristique. Cette propriété est énoncée dans le théorème suivant.

Théorème 3.6 Représentation de Lévy-Khintchine

La fonction caractéristique d'une variable aléatoire Z dont la distribution F est infiniment divisible est de la forme

$$\phi_Z(u) = \exp \left\{ ibu - \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}-\{0\}} (e^{iux} - 1 - iux \cdot 1_{\{|x|<1\}}) \nu(dx) \right\}, \quad (3.3)$$

où le triplet $(b, \sigma, \nu(\cdot)) \in \mathbb{R} \times \mathbb{R}_+ \times (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est composé des **caractéristiques** de Z . La fonction borélienne ν est appelée mesure de Lévy (ou spectre de F) et vérifie les propriétés

suivantes

$$\nu(\mathbb{R} \setminus]-\epsilon, \epsilon[) < \infty, \quad \forall \epsilon > 0; \quad (3.4)$$

$$\nu(\{0\}) = 0; \quad (3.5)$$

$$\int_{|x| \leq 1} x^2 \nu(\mathrm{d}x) < \infty. \quad (3.6)$$

Réciproquement, toute fonction caractéristique prenant la forme (3.3) correspond à une fonction de distribution infiniment divisible.

Exemple 3.1.2

1. Dans le cas d'une loi normale $\mathcal{N}(\mu, \sigma^2)$, les caractéristiques sont ($b = \mu, \sigma = \sigma, \nu(\mathrm{d}x) \equiv 0$).
2. Dans le cas d'une loi Poisson composée, les caractéristiques sont (en reprenant les notations de l'Exemple 3.1.1) ($b = 0, \sigma = 0, \nu(\mathrm{d}x) = \lambda F_X(\mathrm{d}x)$).

On peut voir ce problème sous un autre angle. Imaginons qu'au lieu de chercher à partir d'une variable aléatoire à savoir si elle peut être divisée en une infinité de parties se comportant à l'identique, on se demande comment se comporte une somme infinie centrée réduite de variables aléatoires indépendantes. On constate alors que ces distributions infiniment divisibles présente un attrait supplémentaire. En effet, puisque elles correspondent aux distributions de convolutions asymptotiques, elles sont par conséquent étroitement liées à d'importants résultats comme le Théorème Central Limite et sa version généralisée.

3.2 Théorème Central Limite généralisé et domaines d'attraction

Il existe de nombreuses versions du Théorème Central Limite. Un traitement exhaustif et par conséquent les versions les plus sophistiquées du sujet peuvent être trouvées dans Feller (1971). Pour illustration du propos, la version la plus simple, suffisante, est présentée.

Théorème 3.7 Théorème Central Limite

Soit X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées, d'espérance nulle et de variance unitaire. Dans ce cas,

$$\frac{X_1 + \dots + X_n}{\sqrt{n}}$$

converge en distribution vers une loi normale centrée réduite.

Le Théorème 3.7 peut être vu sous un autre angle. Pour cela, définissons dans un cadre un peu plus général la notion de domaine d'attraction.

Définition 3.8 Soit X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées selon une loi F . Si la suite des sommes partielles réduites

$$S_n^* = \frac{\sum_{i=1}^n X_i}{b_n} - a_n, \quad (3.7)$$

où $(a_n, n \in \mathbb{N})$ est une suite de nombres réels et $(b_n, n \in \mathbb{N})$ une suite de nombres réels positifs, possède une distribution asymptotique G , la distribution F est dite attirée par G . Le **domaine d'attraction** de G est composé de toutes les distributions attirées par G .

Au vu de cette définition, on peut effectuer une relecture du Théorème 3.7 et en conclure que toutes les distributions dont le deuxième moment est fini font partie du domaine d'attraction de la loi normale. A partir de là, deux questions apparaissent. D'une part, quelles distributions représentent un pôle et possèdent un domaine d'attraction et d'autre part, qu'advient-il des distributions dont le second moment n'est pas fini? Font-elles tout de même partie d'un domaine d'attraction? Pour répondre à la première question, il est nécessaire d'introduire la notion de stabilité d'une distribution.

Définition 3.9 Soit X, X_1, \dots, X_n des variables aléatoires indépendantes et identiquement distribuées selon une distribution G et soit $S_n = \sum_{i=1}^n X_i$. La distribution G est dite stable si pour tout $n \in \mathbb{N}$, il existe des constantes c_n et a_n telles que

$$S_n \stackrel{D}{=} c_n X + a_n \quad (3.8)$$

et qu'elle n'est pas concentrée en un seul point. La distribution G est dite strictement stable si (3.8) est vérifiée avec $a_n = 0$.

En particulier, l'égalité en distribution (3.8) établit qu'une distribution stable est également infiniment divisible. La différence majeure étant que dans l'équation (3.2) qui définit une loi infiniment divisible, il n'est pas nécessaire que pour tout $n \in \mathbb{N}$, les n variables aléatoires indépendantes et identiquement distribuées possèdent la même loi que la variable considérée. Lévy (1924, p.69) a montré le résultat suivant.

Théorème 3.10 Soit S_n^* une somme partielle réduite donnée par (3.7). La condition nécessaire pour que la distribution G soit la distribution asymptotique de S_n^* est que G soit stable.

Le Théorème 3.10 n'est pas un théorème d'existence. Il précise simplement que si la loi limite doit exister, elle sera forcément stable. Les différentes versions du Théorème Central Limite assure que pour autant que les seconds moments des variables X_1, \dots, X_n soient finis, elles sont attirées par la loi normale. Il faut donc se demander si parmi les lois dont le second moment est infini, il en existe tout de même certaine qui font partie d'un domaine d'attraction. Le Théorème Central Limite généralisé suivant remplace l'hypothèse d'un second moment fini par des conditions de régularité des queues. Ainsi, il caractérise les distributions telles que les sommes partielles réduites données par (3.7) admettent une loi limite, forcément stable au vu du Théorème 3.10. Pour pouvoir énoncer ce théorème, la définition suivante est encore nécessaire.

Définition 3.11 Une fonction positive H définie sur \mathbb{R}_+ varie doucement si pour $x > 0$,

$$\lim_{s \rightarrow \infty} \frac{H(sx)}{H(s)} \rightarrow 1.$$

Théorème 3.12 Théorème Central Limite généralisé

Une distribution F appartient au domaine d'attraction d'une distribution G si et seulement si il existe une fonction H qui varie doucement telle que

$$\begin{aligned} U(x) &= \int_{-x}^x y^2 F(dy) \sim x^{2-\alpha} H(x), \quad 0 < \alpha \leq 2; \\ \lim_{x \rightarrow \infty} \frac{1 - F(x)}{1 - F(x) + F(-x)} &\rightarrow p; \\ \lim_{x \rightarrow \infty} \frac{F(-x)}{1 - F(x) + F(-x)} &\rightarrow q; \end{aligned}$$

La fonction $U(\cdot)$ fournit en fait une version tronquée du second moment de la loi F . La première condition n'interdit pas un second moment infini, mais impose un type de régularité sur la manière avec laquelle la fonction $U(\cdot)$ atteint son supremum. Les deux autres conditions concernent chacune individuellement une queue de la distribution. Ce sont là également des règles de contrôle sur l'évolution asymptotique de la distribution. Si la distribution est symétrique, les limites sont égales.

Par conséquent, des distributions autres que la loi normale peuvent être considérées sans pour autant remettre en cause une certaine convergence. Cette constatation permet donc de modéliser des sources aléatoires, dont les incréments ou les infimes parties ne possèdent pas un comportement homogène. Dans cet environnement, il a été établi par le Théorème 3.10 entre autre que les lois stables y jouent un rôle charnier, au même titre que la loi normale pour les lois dont le second moment est fini. La prochaine section étudie plus en détail les propriétés de ces distributions.

3.3 Les distributions stables

3.3.1 Description générale

La distribution stable nécessite quatre paramètres pour être complètement décrite. On utilisera par la suite la notation $X \sim S_k(\alpha, \beta, \sigma, \mu)$, où k est un indice qui stipule quelle paramétrisation est utilisée. Le premier paramètre, $\alpha \in]0, 2]$ (on appelle également ces distributions, les distributions α -stables), est dit **indice de stabilité**. En fait, il mesure la rapidité avec laquelle les queues de la distributions se couchent. Si $\alpha = 2$, c'est le cas gaussien et les queues adoptent un comportement exponentiel. Si $\alpha < 2$, les queues adoptent un comportement en loi de puissance. Autrement dit,

$$\lim_{x \rightarrow \infty} x^\alpha \mathbf{P}(X > x) = \lim_{x \rightarrow \infty} x^\alpha \mathbf{P}(X < -x) = \text{cste}(\alpha). \quad (3.9)$$

Les autres exposants sont $\beta \in [-1, 1]$, un **coefficient d'asymétrie**, $\sigma > 0$, un **paramètre d'échelle** et $\mu \in \mathbb{R}$, un **paramètre de lieu**. La représentation de Lévy-Khintchine, donnée par (3.3), fournit le résultat suivant dans le cas d'une variable aléatoire stable.

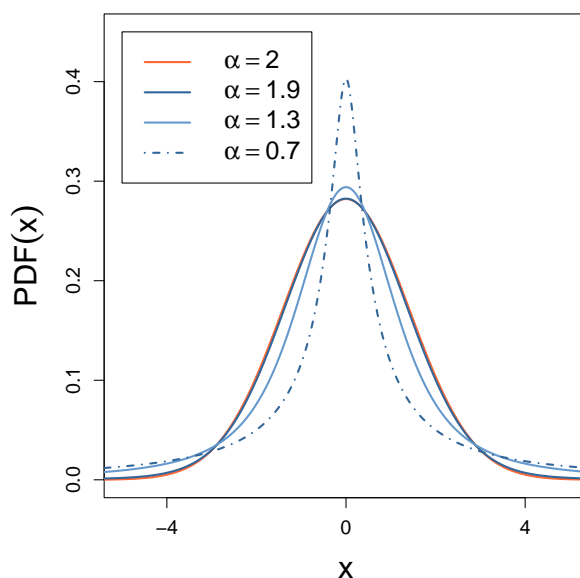


Figure 3.1. Représentation de densités stables selon la paramétrisation donnée par (3.11), pour $\beta = 0$, $\sigma = 1$, $\mu_o = 0$ et $\alpha = 0.7, 1.3, 1.9, 2$. La forme globale est celle d'une cloche (unimodale). À mesure qu' α décroît, le pic devient de plus en plus grand, la région couverte par le pic se réduit et les queues s'alourdissent. Comme le coefficient β est nul, les densités sont symétriques. La différence entre $\alpha = 2$ (loi normale) et $\alpha = 1.99$ peut paraître faible et pourtant elle suffit à détruire toute structure du second moment.

3.3.2 Fonction caractéristique

Théorème 3.13 Si $X \sim S(\alpha, \beta, \sigma, \mu)$, alors ses caractéristiques, définies dans le Théorème 3.6, sont de la forme :

1. $(\mu, \sigma^2, \nu(dx) \equiv 0)$, si $\alpha = 2$, $X \sim \mathcal{N}(\mu, \sigma^2)$;
2. $(\mu, 0, \nu(dx) = (c_1/|x|^{\alpha+1}) \cdot 1_{\{x < 0\}} dx + (c_2/x^{\alpha+1}) \cdot 1_{\{x > 0\}} dx)$, si $\alpha < 2$, où $c_1, c_2 \geq 0$ et $c_1 + c_2 > 0$. En fait, $\beta = (c_2 - c_1)/(c_2 + c_1)$.

Néanmoins, ce sont d'autres formulations paramétrisées de la fonction caractéristique qui sont le plus souvent employées, spécialement en statistique. La première est donnée dans le théorème ci-dessous.

Théorème 3.14 La fonction caractéristique d'une variable aléatoire stable $S_1(\alpha, \beta, \sigma, \mu_1)$ est donnée par

$$\phi_1(u) = \begin{cases} \exp \{i\mu_1 u - \sigma^\alpha |u|^\alpha (1 - i\beta \text{signe}(u) \tan(\pi\alpha/2))\} , & \text{si } \alpha \neq 1 ; \\ \exp \{i\mu_1 u - \sigma |u| (1 + i\beta(2/\pi) \text{signe}(u) \log |u|)\} , & \text{si } \alpha = 1 . \end{cases} \quad (3.10)$$

Réciproquement, si une variable aléatoire X possède la fonction caractéristique (3.10), alors $X \sim S_1(\alpha, \beta, \sigma, \mu_1)$.

Cette paramétrisation possède une faiblesse. Elle est discontinue en $\alpha = 1$. Cependant une réécriture due à Zolotarev (1986) permet de dépasser cette difficulté.

$$\phi_0(u) = \begin{cases} \exp \{i\mu_0 u - \sigma^\alpha |u|^\alpha (1 + i\beta \text{signe}(u) \tan(\pi\alpha/2)(|\sigma u|^{1-\alpha} - 1))\} , & \text{si } \alpha \neq 1 ; \\ \exp \{i\mu_0 u - \sigma |u| (1 + i\beta(2/\pi) \text{signe}(u) \log(\sigma |u|))\} , & \text{si } \alpha = 1 . \end{cases} \quad (3.11)$$

On notera, dans ce cas, $X \sim S_0(\alpha, \beta, \sigma, \mu_0)$. La fonction caractéristique donnée par (3.11) est une fonction continue en ses quatre paramètres. La correspondance entre les paramètres μ_1 apparaissant dans (3.10) et μ_0 dans (3.11) est donnée par

$$\mu_0 = \begin{cases} \mu_1 + \beta\sigma \tan(\pi\alpha/2) & \text{si } \alpha \neq 1 , \\ \mu_1 + \beta(2/\pi)\sigma \log \sigma & \text{si } \alpha = 1 . \end{cases} \quad (3.12)$$

3.3.3 Propriétés d'une loi stable

Les propriétés qui sont énoncées ci-dessous ne le sont que pour la paramétrisation $S_0(\alpha, \beta, \sigma, \mu_0)$. Elles sont également vérifiées pour la paramétrisation $S_1(\alpha, \beta, \sigma, \mu_1)$, même si les termes dans les égalités peuvent légèrement varier. Un bon résumé est fourni par Nolan (1998). Les premières propriétés présentées ont trait aux paramètres de la loi stable.

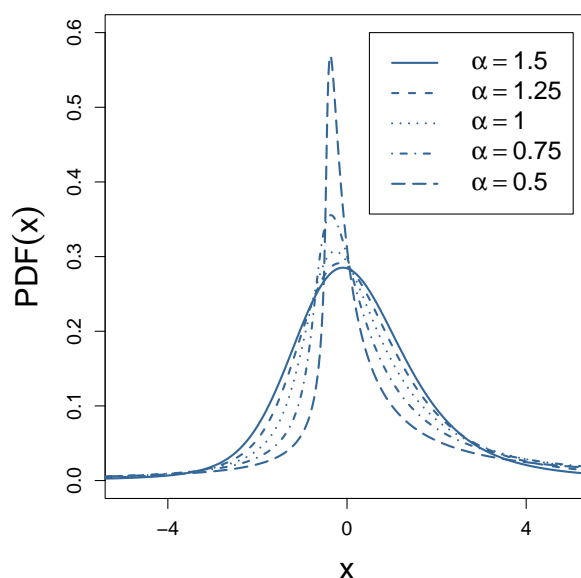


Figure 3.2. Représentation de densités stables selon la paramétrisation donnée par (3.11), pour $\beta = 0.5$, $\sigma = 1$, $\mu_o = 0$ et $\alpha = 0.5, 0.75, 1, 1.25, 1.5$. Comme le coefficient β est non-nul, les densités sont asymétriques. Comme il est positif, la masse se déplace vers la droite à mesure que α décroît.

L'Equation (3.9) a présenté le rôle du paramètre α , à savoir que les queues se comportent asymptotiquement comme une loi de puissance et que plus α est grand, plus les queues sont minces. Concernant le paramètre β , on peut présenter la propriété suivante qui emphase la relation entre β et la notion de symétrie de la distribution.

Propriété 3.15 Réflexion

Si $X \sim S_0(\alpha, \beta, 1, 0)$ et $Y \sim S_0(\alpha, -\beta, 1, 0)$, alors $X = -Y$ et $F_Y(y) = 1 - F_X(y)$.

Par conséquent, si $\beta = 0$, la loi est symétrique. La propriété suivante à trait aux paramètres d'échelle et de lieu et fourni la relation entre une variable stable quelconque et une variable stable standardisée.

Propriété 3.16 Standardisation

Si $X \sim S_0(\alpha, \beta, \sigma, \mu_0)$, alors

$$Z = \begin{cases} (X - \mu_0)/\sigma + \beta \tan(\pi\alpha/2) & \text{si } \alpha \neq 1, \\ (X - \mu_0)/\sigma & \text{si } \alpha = 1, \end{cases} \quad (3.13)$$

suit une distribution $S_0(\alpha, \beta, 1, 0)$.

Propriété 3.17 Moments

Si $X \sim S_0(\alpha, \beta, \sigma, \mu_0)$, alors $E(|X|^l) < \infty$ si et seulement si $0 < l < \alpha$.

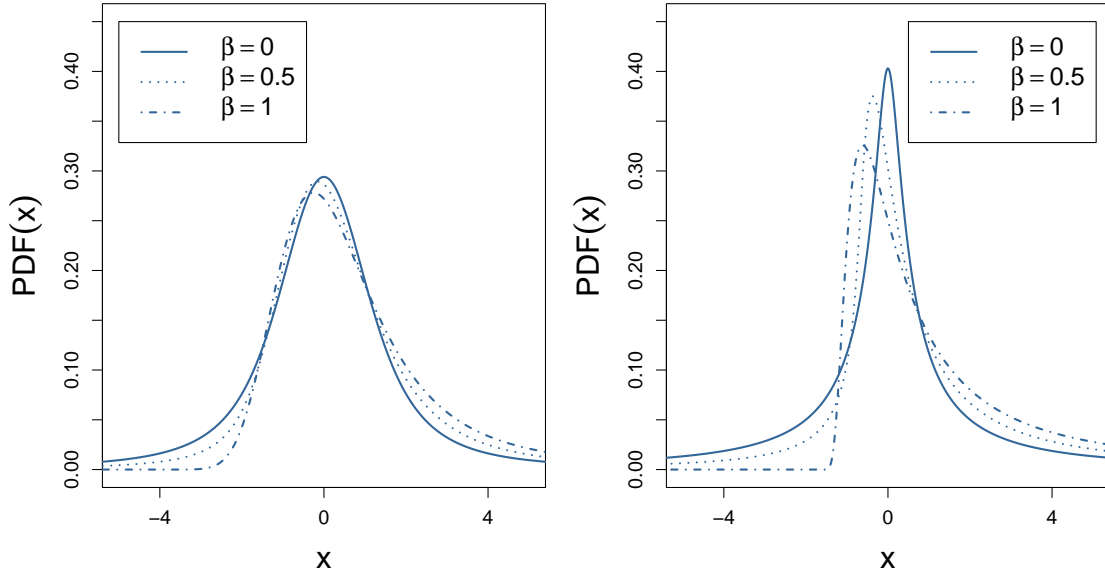


Figure 3.3. Représentation de densités stables selon la paramétrisation donnée par (3.11). A gauche, $\alpha = 1.3$ et à droite $\alpha = 0.7$. Le paramètre β prend les valeurs 0, 0.5, 1. L'effet d'asymétrie est d'autant plus fort que α est faible.

Cette propriété découle du comportement en loi de puissance de la distribution. En effet, le calcul du $r^{\text{ème}}$ moment nécessite le calcul d'intégrale du type

$$\int_c^\infty x^r f_X(x) dx \propto \int_c^\infty x^{r-\alpha-1} dx,$$

où $c > 0$ est un nombre positif arbitrairement grand. Cette intégrale converge si et seulement si $0 < r < \alpha$. On en déduit que seule la distribution gaussienne possède un second moment (et donc une variance) fini et qu'une variable aléatoire stable possède une espérance finie si et seulement si $\alpha > 1$. Dans ce cas, si $X \sim S_0(\alpha, \beta, \sigma, \mu_0)$, alors

$$E(X) = \mu_0 - \beta\sigma \tan(\pi\alpha/2). \quad (3.14)$$

Il faut remarquer que dans la paramétrisation $S_1(\alpha, \beta, \sigma, \mu_1)$, l'espérance coïncide avec le paramètre μ_1 . Par définition, il existe des coefficients tels que la combinaison linéaire de deux variables aléatoires stables est encore stable. Les deux propriétés suivantes explicitent ce fait en termes des paramètres des lois considérées.

Propriété 3.18 Transformation linéaire

Si $X \sim S_0(\alpha, \beta, \sigma, \mu_0)$ et $a, b \in \mathbb{R}$, $a \neq 0$, alors

$$aX + b \sim S_0(\alpha, \beta \text{signe}(a), |a|\sigma, a\mu_0 + b).$$

Propriété 3.19 Combinaison linéaire

Si $X_\star \sim S_0(\alpha, \beta_\star, \sigma_\star, \mu_\star)$, $X_\dagger \sim S_0(\alpha, \beta_\dagger, \sigma_\dagger, \mu_\dagger)$ et $X_\star \perp X_\dagger$, alors $X_\star + X_\dagger \sim S_0(\alpha, \beta, \sigma, \mu_0)$, où

$$\begin{aligned}\beta &= (\beta_\star \sigma_\star^\alpha + \beta_\dagger \sigma_\dagger^\alpha) / (\sigma_\star^\alpha + \sigma_\dagger^\alpha), \\ \sigma &= \sqrt[\alpha]{\sigma_\star^\alpha + \sigma_\dagger^\alpha}, \\ \mu_0 &= \begin{cases} \mu_\star + \mu_\dagger + \tan(\pi\alpha/2)(\beta\sigma - \beta_\star\sigma_\star - \beta_\dagger\sigma_\dagger) & \text{si } \alpha \neq 1, \\ \mu_\star + \mu_\dagger + (2/\pi)(\beta\sigma \log \sigma - \beta_\star\sigma_\star \log \sigma_\star - \beta_\dagger\sigma_\dagger \log \sigma_\dagger) & \text{si } \alpha = 1. \end{cases}\end{aligned}$$

3.3.4 Fonctions de densité

Les lois stables ne possèdent, en général, pas de formulation explicite de leur fonction de distribution, ni de leur fonction de densité. Ces fonctions possèdent des propriétés qui méritent tout de même d'être mentionnées.

Propriété 3.20 Continuité

Toute loi stable possède une fonction de densité continue et infiniment différentiable.

Propriété 3.21 Mode

Toute loi stable est unimodale (Yamazato, 1978).

Propriété 3.22 Support

Si $X \sim S_0(\alpha, \beta, \sigma, \mu_0)$, le support de sa densité est donné par

$$\begin{cases}]\mu_0 - \sigma \tan(\pi\alpha/2), +\infty[, & \text{si } \alpha < 1 \text{ et } \beta = 1, \\]-\infty, \mu_0 + \sigma \tan(\pi\alpha/2)[, & \text{si } \alpha < 1 \text{ et } \beta = -1, \\]-\infty, +\infty[& \text{sinon.} \end{cases}$$

Il existe trois situations pour lesquelles une formulation explicite de la fonction de densité et de la fonction de distribution existent. Ces cas sont énumérés ci-après.

Loi Normale : lorsque $\alpha = 2$. La loi stable coïncide avec la loi normale d'espérance $\mu = \mu_0$ et de variance $\tilde{\sigma}^2 = 2\sigma^2$. La fonction caractéristique s'écrit :

$$\phi(u) = \exp\{i\mu_0 u - \sigma|u|\}.$$

Loi de Cauchy : lorsque $\alpha = 1$ et $\beta = 0$. La loi stable coïncide avec la loi de Cauchy de paramètre de lieu μ_0 et de paramètre d'échelle σ . La fonction caractéristique s'écrit :

$$\phi(u) = \exp\{i\mu_0 u - \sigma^2 u^2\}.$$

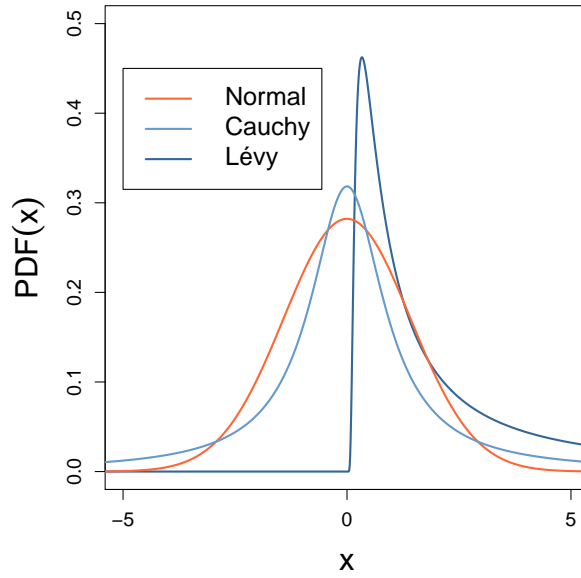


Figure 3.4. Représentation des densités d'une loi normale $\mathcal{N}(\mu_0 = 0, \tilde{\sigma}^2 = 2\sigma^2 = 2)$, d'une loi de Cauchy($\mu_0 = 0, \sigma = 1$) et d'une loi de Lévy($\mu_1 = 0, \sigma = 1$). Ces trois densités correspondent aux cas particuliers pour lesquelles une version explicite de la densité d'une loi stable existe.

Loi de Lévy : lorsque $\alpha = 1/2$ et $\beta = \pm 1$. La loi stable coïncide avec la loi de Lévy de paramètre de lieu μ_1 et de paramètre d'échelle σ . La fonction caractéristique s'écrit, dans le cas $\beta = 1$:

$$\begin{aligned}\phi(u) &= \exp \left\{ i\mu_0 u - \sqrt{\sigma|u|} \left(1 + i \operatorname{signe}(u) (\sqrt{\sigma|u|} - 1) \right) \right\} \\ &= \exp \{ i\mu_1 u - \sqrt{\sigma|u|} (1 - i \operatorname{signe}(u)) \} .\end{aligned}$$

3.3.5 Simulation

Dans le cas général, le résultat suivant proposé par Chambers *et al.* (1976) fournit une méthode pour simuler n'importe quelle variable aléatoire stable.

Théorème 3.23 Soit U une variable aléatoire uniforme $\mathcal{U}[-\pi/2, \pi/2]$ et W une variable aléatoire exponentielle $\mathcal{E}(\lambda = 1)$. Pour $0 < \alpha \leq 2$ et $\xi = \arctan(\beta \tan(\pi\alpha/2))/\alpha$, la variable aléatoire

$$X = \begin{cases} \frac{\sin(\alpha(\xi + U))}{(\cos(\alpha\xi) \cos(U))^{1/\alpha}} \left[\frac{\cos(\alpha\xi + (\alpha - 1)U)}{W} \right]^{(1-\alpha)/\alpha} & \text{si } \alpha \neq 1, \\ \frac{2}{\pi} \left[\left(\frac{\pi}{2} + \beta U \right) \tan U - \beta \log \left(\frac{\frac{\pi}{2} W \cos U}{\frac{\pi}{2} + \beta U} \right) \right] & \text{si } \alpha = 1, \end{cases}$$

possède une distribution $S_0(\alpha, \beta, 1, 0)$. Pour générer une variable aléatoire avec un paramètre de lieu μ_0 et un paramètre d'échelle σ arbitraires, il suffit d'appliquer le résultat donné par (3.13).

Il existe des cas particuliers où la méthode est simple. En fait, ces cas correspondent aux cas particuliers pour lesquels une forme explicite de la densité existe et qui sont présentés dans la Section 3.3.4.

Loi Normale : l'algorithme de Box-Muller peut s'appliquer. Si $U_1, U_2 \sim \mathcal{U}[0, 1]$, alors

$$\begin{aligned} X_1 &= \mu + \sigma \sqrt{-2 \log U_1} \cos(2\pi U_2), \\ X_2 &= \mu + \sigma \sqrt{-2 \log U_1} \sin(2\pi U_2), \end{aligned}$$

sont deux variables aléatoires $\mathcal{N}(\mu, \sigma^2)$ indépendantes.

Loi de Cauchy : la transformation

$$X = \sigma \tan(\pi(U - 1/2)) + \mu,$$

où $U \sim \mathcal{U}[0, 1]$, fournit une variable aléatoire Cauchy(μ, σ).

Loi de Lévy : la transformation

$$X = \sigma \frac{1}{Z^2} + \mu,$$

où $Z \sim \mathcal{N}(0, 1)$, fournit une variable aléatoire Lévy(μ, σ).

3.3.6 Estimation

L'estimation des paramètres d'une loi stable est rendue difficile par l'absence d'une forme explicite de la fonction de densité (à l'exception des cas présentés dans la Section 3.3.4). On peut grossièrement subdiviser les approches en trois groupes. Les méthodes basées sur les quantiles empiriques ; celles basées sur la fonction caractéristique empirique et celles basées sur le maximum de vraisemblance. L'ordre de présentation des méthodes correspond également à celui de leur développement.

Méthodes des quantiles

Fama and Roll (1971) ont proposé un estimateur pour le cas centré, symétrique. Leur méthode souffre principalement de la limitation sur les valeurs possibles pour les paramètres α et β . McCulloch (1986) a généralisé cette approche basée sur les quantiles et fournit des estimateurs consistents pour les quatre paramètres, pour autant que $\alpha \geq 0.6$.

Il a défini les fonctions :

$$\eta_\alpha = \frac{x(0.95) - x(0.05)}{x(0.75) - x(0.25)} \quad \text{et} \quad \eta_\beta = \frac{x(0.95) - 2x(0.50) + x(0.05)}{x(0.95) - x(0.05)}, \quad (3.15)$$

fonctions indépendantes des paramètres de lieu et d'échelle. Par conséquent, la relation entre les paramètres α et β et ces fonctions peut être inversée et les paramètres estimés grâce aux valeurs empiriques des fonctions (3.15) et une interpolation des tables fournies par l'auteur. Les deux autres paramètres sont estimés de manière similaire.

Méthodes de la fonction caractéristique

Press (1972) a proposé une méthode basée sur des transformations de la fonction caractéristique. Les paramètres sont également estimés par paires. Tout d'abord, pour tout $\alpha \in (0, 2]$, on peut écrire :

$$|\phi(u)| = \exp(\sigma^\alpha \cdot |u|^\alpha). \quad (3.16)$$

Par conséquent, en considérant deux valeurs distinctes $u_1 \neq u_2$ et en inversant la relation (3.16), on obtient les estimateurs

$$\hat{\alpha} = \log \frac{\log |\hat{\phi}(u_1)|}{\log |\hat{\phi}(u_2)|} \bigg/ \log \left| \frac{u_1}{u_2} \right|$$

et

$$\log \hat{\sigma} = \left(\log |u_1| \log(-\log |\hat{\phi}(u_2)|) - \log |u_2| \log(-\log |\hat{\phi}(u_1)|) \right) \bigg/ \log \left| \frac{u_1}{u_2} \right|,$$

où $\hat{\phi}(u) = n^{-1} \sum_{j=1}^n \exp(iux_j)$. Pour estimer les deux paramètres restants, une technique similaire est appliquée à la partie imaginaire du logarithme de la fonction caractéristique en s'aidant d'une autre paire de points distincts $u_3 \neq u_4$. Les estimateurs dérivent leur consistance du fait que l'estimateur de la fonction caractéristique est consistant. Le choix des quatre points pour lesquels est estimée la fonction caractéristique est problématique. Koutrouvelis (1980) a proposé le choix $u_1 = 0.2$, $u_2 = 0.8$, $u_3 = 0.1$ et $u_4 = 0.4$.

Koutrouvelis (1981) a proposé une méthode basée sur l'équation suivante, dérivée de (3.11),

$$\log(-\log |\phi(u)|^2) = \log(2\sigma^\alpha) + \alpha \log |u|. \quad (3.17)$$

Puisque cette équation ne dépend que des paramètres α et σ , l'auteur propose d'estimer ces paramètres en régressant la variable réponse $\log(-\log |\phi(u_j)|^2)$ sur la variable explicative $\log |u_j|$ pour des valeurs réelles u_j , $j = 1, \dots, J$. Cette première régression fournit

les estimations $\hat{\alpha}$ et $\hat{\sigma}$ qui, une fois injectée dans

$$\arctan \left(\frac{\operatorname{Im}(\phi(u))}{\operatorname{Re}(\phi(u))} \right) = \mu u + \beta \sigma^\alpha \tan \left(\frac{\pi \alpha}{2} \right) \operatorname{signe}(u) |u|^\alpha, \quad (3.18)$$

permet d'effectuer une seconde régression pour estimer les deux autres paramètres. La procédure est appliquée de manière itérative jusqu'à ce qu'un critère d'arrêt soit satisfait. Kogon and Williams (1998) ont adapté cette dernière méthode en supprimant le côté itératif et en simplifiant la méthode de régression. Leur méthode est par conséquent plus rapide, mais également moins précise que celle proposée par Koutrouvelis (1981).

Méthodes du maximum de vraisemblance

La méthode du maximum de vraisemblance n'est pas fréquemment appliquée à ce problème étant donné sa complexité résultant de l'absence de version explicite de la fonction de densité. Pourtant, cette méthode possède des qualités indéniables, la première étant qu'elle délivre également des informations sur la distribution de l'estimateur (Dumouchel, 1973). Les deux méthodes les plus fréquemment citées sont dues à Nolan (2001), qui procède par intégration numérique directe et à Mittnik *et al.* (1999) qui ont proposé une estimation qui procède à l'optimisation sous contrainte par l'intermédiaire de la méthode des transformées de Fourier rapides.

Pour conclure cette section, il faut encore signaler que ces méthodes fonctionnent toutes relativement bien pour autant que le jeu de données soumis provienne bel et bien d'une loi stable. Un test formel pour l'attester n'existant pas, une inspection visuelle et une estimation de l'indice de stabilité de la distribution est souhaitable.

3.4 Drap stable

Comme le mentionne Feller (1971, p.557), l'intérêt des distributions infiniment divisibles (et en particulier les distributions stables) apparaît lorsque l'on traite de processus stochastiques stationnaires à accroissements indépendants. Par exemple, lorsque $d = 1$, pour toute partition $t_0 < t_1 < \dots < t_n$, les accroissements $Z(t_k) - Z(t_{k-1})$ représentent n variables aléatoires indépendantes. La stationnarité du processus signifie que la distribution de $Z(t + s) - Z(t)$ ne dépend pas de t mais seulement de la taille de l'écart considéré s . Autrement dit $Z(t + s) - Z(t)$ peut être vu comme la somme de n variables aléatoires indépendantes distribuées comme $Z(t + s/n) - Z(t)$. La loi des accroissements d'un tel processus est donc bien infiniment divisible. Il en va de même lorsque $d = 2$, c'est-à-dire lorsque l'on s'intéresse aux processus stochastiques stationnaires à accroissement indépendants en deux dimensions, appelés également draps stochastiques.

Dans cette section, on considère $\{Z(s); s \in \mathbb{R}_+^2\}$, un processus stochastique défini sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$ donné. Pour deux points donnés, s et $t \in \mathbb{R}_+^2$, on écrit $s \leq t$ si $s_1 \leq t_1$ et $s_2 \leq t_2$. Si $s \leq t$, on dénote par $R = (s, t]$ le rectangle inclus dans \mathbb{R}_+^2 dont les sommets sont $s = (s_1, s_2)$, (s_1, t_2) , (t_1, s_2) et $t = (t_1, t_2)$. L'accroissement du processus Z sur R est défini comme

$$\Delta_R Z = Z(t_1, t_2) - Z(s_1, t_2) - Z(t_1, s_2) + Z(s_1, s_2).$$

Définition 3.24 Le processus stochastique $\{Z(s); s \in \mathbb{R}_+^2\}$ possède des **accroissements indépendants** si pour tout ensemble fini de rectangles deux-à-deux disjoints R_1, \dots, R_n , les accroissements $\Delta_{R_1}, \dots, \Delta_{R_n}$ sont des variables aléatoires indépendantes.

Si, sur la base de $\{Z(s); s \in \mathbb{R}_+^2\}$, on définit un nouveau processus stochastique $\{\tilde{Z}(s); s \in \mathbb{R}_+^2\}$ comme $\tilde{Z}(s) = Z(s) - Z(s_1, 0) - Z(0, s_2) + Z(0, 0)$, on remarque alors que $\tilde{Z} = 0$ sur les axes et que $\Delta_R \tilde{Z} = \Delta_R Z$ pour tout $R \in \mathbb{R}_+^2$. Par conséquent, il n'y a aucune perte de généralité à supposer que le processus sous étude $\{Z(s); s \in \mathbb{R}_+^2\}$ est identiquement nul sur les axes. Selon Adler *et al.* (1983), cette propriété est d'ailleurs requise dans la définition même d'un processus stochastique additif.

Définition 3.25 Un processus stochastique à accroissements indépendants $\{Z(s); s \in \mathbb{R}_+^2\}$, tel que la fonction $s \longrightarrow Z(s)$ est continue en probabilité, est appelé **drap de Lévy**.

Puisque les incréments du processus $Z(\cdot)$ sont indépendants, on peut montrer que l'hypothèse de continuité en probabilité de $Z(\cdot)$ revient à la continuité presque sûrement de $Z(\cdot)$ pour tout $s \in \mathbb{R}_+^2$. En particulier, $Z(s)$ possède une distribution infiniment divisible pour tout $s \in \mathbb{R}_+^2$. Par conséquent, le Théorème 3.6 stipule que pour tout $s \in \mathbb{R}_+^2$, il existe des caractéristiques $(b_s, \sigma_s, \nu_s(\cdot))$ telles que la fonction caractéristique de $Z(s)$ s'écrit comme (3.3). Si l'on considère R_1, \dots, R_m , m rectangles disjoints, les variables aléatoires associées sont indépendantes et par conséquent, le logarithme de la fonction caractéristique associée à la somme de ces variables aléatoires est égal à la somme des logarithmes des fonctions caractéristiques de chacune des variables aléatoires. On en conclut que les termes du triplet $(b_s, \sigma_s, \nu_s(\cdot))$ se doivent d'être des mesures additives relativement à $\mathcal{B}(\mathbb{R}_+^2)$, l'ensemble des boréliens de \mathbb{R}_+^2 . On peut donc réécrire la formule de Lévy-Khintchine pour un tel processus.

Théorème 3.26 Soit $\{Z(s); s \in \mathbb{R}_+^2\}$ un drap de Lévy univarié défini sur \mathbb{R}_+^2 . Dans ce cas, on a que

$$\begin{aligned}\phi_{Z(s)}(u) &= \mathbb{E}(\exp(iuZ(s))) , \\ &= \exp \left\{ iub(s) - \frac{1}{2}\sigma(s)u^2 + \int_{\mathbb{R}-\{0\}} (e^{iux} - 1 - iux \cdot 1_{\{|x|<1\}})\nu(ds, dx) \right\} , \quad (3.19)\end{aligned}$$

où le triplet $(b(s), \sigma(s), \nu(\cdot, \cdot))$ est tel que

- l'application $s \rightarrow b(s)$ est continue à valeur réelle, identiquement nulle pour tout $s \in \mathbb{R}_+^2$ tel que $s_1 = 0$ ou $s_2 = 0$;
- l'application $\mathcal{B}(\mathbb{R}_+^2) \rightarrow \sigma^2(\mathcal{B}(\mathbb{R}_+^2))$ est une mesure additive ;
- l'application $\nu(ds, dx)$ est une mesure de Lévy généralisée. En outre, elle vérifie que $\nu(A, \{0\}) = 0$ pour tout $A \in \mathcal{B}(\mathbb{R}_+^2)$.

Réciproquement, pour tout triplet $(b(s), \sigma(s), \nu(\cdot, \cdot))$, il existe un processus de Lévy, $Z(\cdot)$ tel que (3.19) soit vérifiée.

On peut encore rajouter les deux définitions suivantes qui permettent de caractériser un processus de Lévy.

Définition 3.27 Le processus stochastique $\{Z(s); s \in \mathbb{R}_+^2\}$ est dit **centré** si pour tout $s \in \mathbb{R}_+^2$, la composante de tendance $b(s)$ est identiquement nulle.

Définition 3.28 Un processus stochastique $\{Z(s); s \in \mathbb{R}_+^2\}$ qui est tel que la distribution de ses incréments $\Delta_R Z = Z(t_1, t_2) - Z(s_1, t_2) - Z(t_1, s_2) + Z(s_1, s_2)$ ne dépend que de son aire et non de sa localisation est dit **isotropique**.

Une autre manière de lire le résultat du Théorème 3.26 est d'écrire qu'un drap de Lévy peut être vu comme la somme d'une tendance déterministe, d'un drap brownien et d'un processus de saut. Pour mieux constater ce fait-là, il faut tout d'abord définir ce qu'est un drap Brownien ainsi que la mesure aléatoire de Poisson associée à la mesure de Lévy $\nu(\cdot)$ considérée.

3.4.1 Le drap brownien

La première remarque est que comme tout processus stochastique additif, le drap brownien peut contenir un terme de tendance non nul. On se contentera néanmoins de le définir sans cette caractéristique.

Définition 3.29 Soit σ^2 une mesure additive non négative sur \mathbb{R}_+^2 , continue et qui s'annule sur les axes. Dans ce cas, un processus stochastique gaussien à valeurs réelles $\{W(s); s \in$

$\mathbb{R}_+^2\}$, d'espérance nulle et de covariance égale à

$$E(W(s)W(t)) = \sigma^2 \left((0, s] \cap (0, t] \right)$$

est appelé **drap brownien**.

Dalang (2003) fournit en première partie d'article, une excellente présentation de ce processus ainsi que de ses principales propriétés. Pour une preuve de celles énoncées ci-après, le lecteur est renvoyé à cette référence.

Propriété 3.30 – La restriction du drap brownien à une droite verticale ou horizontale est un mouvement brownien dont la vitesse est égale à la racine carrée de la coordonnée fixée.

- Le processus $s \rightarrow W(s, 1-s)$, $0 \leq s \leq 1$, qui représente la restriction du drap au segment reliant les points $(0, 1)$ et $(1, 0)$, est un processus gaussien qui s'annule à ses extrémités et représente en fait un pont brownien.
- Le processus $\tilde{W}(s_1, s_2) = s_1 W(1/s_1, s_2)$ est encore un drap brownien.

Cette parenthèse sur le drap brownien étant terminée, on revient à la représentation d'un drap de Lévy. Comme écrit précédemment, il faut encore définir une mesure.

Définition 3.31 La **mesure aléatoire de Poisson** Π_ν associée à la mesure de Lévy $\nu(\cdot)$ est une application possédant les propriétés suivantes :

- pour tout ensemble borélien $A \in \mathcal{B}(\mathbb{R}_+^2)$, $\Pi_\nu(A)$ est une variable aléatoire de Poisson de paramètre $\nu(A)$;
- si A_1, \dots, A_n sont des ensembles boréliens deux à deux disjoints, les variables aléatoires $\Pi(A_1), \dots, \Pi(A_n)$ sont indépendantes.

Puisque Π_ν ne prend que des valeurs entières, cette mesure est une somme de points de masse unitaire. On peut remarquer que si l'on s'intéresse aux points plutôt qu'au masses, on parle d'un processus ponctuel plutôt que d'une mesure aléatoire de Poisson. On peut maintenant énoncer le théorème suivant dont une preuve peut être trouvée dans Dalang and Walsh (1992).

Proposition 3.32 Soit $\{Z(s) : s \in \mathbb{R}_+^2\}$ un drap de Lévy dont la fonction caractéristique est donnée par (3.19). Dans ce cas, il existe des processus $Z^G(\cdot)$ et $Z^J(\cdot)$ et une mesure aléatoire de Poisson Π , tous mesurables par rapport à la tribu $\sigma\{Z(s), s \in \mathbb{R}_+^2\}$ tels que

$$Z(s) = Z^G(s) + Z^J(s), \quad \forall s \in \mathbb{R}_+^2, \text{ presque sûrement,}$$

où $Z^G(\cdot)$ est un processus gaussien indépendant de $Z^J(\cdot)$. Sa fonction caractéristique

s'écrit $\phi_{Z^G}(u) = \exp\{ib(s)u - \frac{1}{2}\sigma^2(s)u^2\}$. Le processus $Z^J(\cdot)$ quant à lui vérifie

$$Z^J(s) = \int_{|z|<1} z \left(\Pi((0, s] \times dz) - \nu((0, s] \times dz) \right) + \int_{|z|\geq 1} z \Pi((0, s] \times dz), \quad \forall s \in \mathbb{R}_+^2.$$

Cette proposition établit que tout drap de Lévy (sans terme déterministe) peut-être vu comme la somme de deux processus indépendants : un drap brownien et un pur processus de sauts associé à une mesure aléatoire de Poisson. On peut énoncer les propriétés suivantes concernant le drap de Lévy présenté tout au long de cette section. Une preuve peut-être trouvée dans Dalang and Walsh (1992).

Propriété 3.33 1. Le processus $\{Z(s) : s \in \mathbb{R}_+^2\}$ est continu à droite et possède des limites dans les autres quadrants ;

2. $\square Z(s) = 0$ sauf pour un ensemble dénombrable de points aléatoires, où

$$\square Z(s) = Z(s) - Z(s_1, s_2-) - Z(s_1-, s_2) + Z(s_1-, s_2-);$$

3. Si $\square Z(s^*) = z$, alors $Z(s_1^*, v) - Z(s_1^*-, v) = z$ pour tout $v \geq s_2^*$ et $Z(u, s_2^*) - Z(u, s_2^*-) = z$ pour tout $u \geq s_1^*$;

4. Les trajectoires de Z n'ont pas d'autres discontinuités que celles présentées dans les deux points précédents ; en particulier,

- l'ensemble $\{s_1 \geq 0 : \exists s_2 \geq 0 : Z(s_1, s_2) \neq Z(s_1-, s_2)\}$ est dénombrable ;
- l'ensemble $\{s_2 \geq 0 : \exists s_1 \geq 0 : Z(s_1, s_2) \neq Z(s_1, s_2-)\}$ est dénombrable ;

5. La probabilité que deux sauts partagent leurs abscisses ou leurs ordonnées est nulle.

Dans ce qui suit, nous allons nous intéresser à des draps de Lévy bien particuliers. Il s'agit des **draps stables isotropiques** ; autrement dit, des draps dont les incréments sont des variables aléatoires stables, $\alpha \in (0, 2]$. Comme pour les variables aléatoires (voir Théorème 3.13), parmi les caractéristiques du processus fournies par (3.19), soit $\alpha = 2$ et la mesure de Lévy généralisée est identiquement nulle, soit $\alpha < 2$ et c'est un processus de sauts pur (*i.e.* $\sigma^2(s)$ est identiquement nulle). Dans le premier cas, on est en présence d'un drap brownien, dans le second, d'un processus de saut pur. Donc, pour revenir à la dernière proposition présentée ci-dessus, la Proposition 3.32, dans le cas qui nous intéresse, on n'observe pas deux processus orthogonaux l'un à l'autre mais soit l'un soit l'autre est identiquement nul. Maintenant que le drap stable isotropique est bien défini, une méthode pour le simuler sur un domaine carré unitaire est présentée, afin de pouvoir l'étudier. Cette méthode représente une généralisation de celle présentée par Dalang and Khoshnevisan (2004) qui eux considèrent des draps stables isotropiques, symétriques et standardisés. La méthode présentée ci-dessous procure un drap stable tel que la distribution de la variable aléatoire associée au point $s = (1, 1)$ soit une distribution $S_0(\alpha, \beta, \sigma, \mu_0)$.

Simulation d'un drap stable de paramètre $(\alpha, \beta, \sigma, \mu_0)$:

1. Dans le carré unitaire, simuler une suite de n^2 variables aléatoires $\{\xi_{i,j}\}_{i,j \geq 1}$ telle que

$$\xi_{ij} \stackrel{\text{iid}}{\sim} S_0(\alpha, \beta, \check{\sigma}, \check{\mu}_0), \quad 1 \leq i, j \leq n,$$

où $\check{\sigma} = n^{-2/\alpha}\sigma$ et $\check{\mu}_0 = n^{-2}\mu_0 - \beta\sigma \tan(\pi\alpha/2)(n^{-2} - n^{-2/\alpha})$.

2. Définir une marche aléatoire comme

$$S_{k,l}^n := \sum_{1 \leq i \leq k} \sum_{1 \leq j \leq l} \xi_{i,j}, \quad 1 \leq k, l \leq n.$$

3. Pour $\alpha \neq 1$, en s'aidant des Propriétés 3.18 et 3.19, on trouve que

$$S_{k,l}^n \sim S_0\left(\alpha, \beta, (kl)^{1/\alpha}\check{\sigma}, (kl)\check{\mu}_0 + \beta\check{\sigma} \tan\left(\frac{\pi\alpha}{2}\right)((kl)^{1/\alpha} - kl)\right).$$

4. Par conséquent, pour $\alpha \neq 1$ et pour tout $s \in [0, 1]^2$, lorsque $n \rightarrow \infty$, on peut écrire

$$S_{[ns_1], [ns_2]}^n \sim S_0\left(\alpha, \beta, (s_1s_2)^{1/\alpha}n^{2/\alpha}\check{\sigma}, (s_1s_2)n^2\check{\mu}_0 + \beta\check{\sigma} \tan\left(\frac{\pi\alpha}{2}\right)((s_1s_2)^{1/\alpha}n^{2/\alpha} - (s_1s_2)n^2)\right).$$

En insérant les valeurs de $\check{\sigma}$ et $\check{\mu}_0$ dans la loi définie ci-dessus, on peut écrire que

$$S_{[ns_1], [ns_2]}^n \sim S_0\left(\alpha, \beta, (s_1s_2)^{1/\alpha}\sigma, (s_1s_2)\mu_0 + \beta\sigma \tan\left(\frac{\pi\alpha}{2}\right)((s_1s_2)^{1/\alpha} - (s_1s_2))\right).$$

Ce résultat peut être réécrit comme

$$S_{[ns_1], [ns_2]}^n \xrightarrow{D} Z(s) \sim (s_1s_2)^{1/\alpha}S_0(\alpha, \beta, \sigma, \tilde{\mu}_0),$$

avec

$$\tilde{\mu}_0 = (s_1s_2)^{1-1/\alpha}\mu_0 + \beta\sigma \tan\left(\frac{\pi\alpha}{2}\right)\left(1 - (s_1s_2)^{1-1/\alpha}\right).$$

Des illustrations de draps obtenus à l'aide de cette méthode sont présentées dans la Figure 3.5. Des quatre paramètres de la loi stable, seul l'indice de stabilité a été varié. Les autres ont été fixés à $\beta = 0$, $\mu = 0$ et $\sigma = \sqrt{2}/2$. La valeur du dernier paramètre assure que lorsque $\alpha = 2$ (en haut, à gauche), on obtient un drap brownien généré à partir de sauts possédant une distribution normale standardisée. Les trois autres valeurs pour α sont, dans l'ordre, 1.6, 1.1 et 0.8. Pour générer ces surfaces, on a pris les valeurs $n = 3000$. Autrement dit, la suite de variables aléatoires indépendantes et identiquement distribuées était composée de neuf millions de sauts et pour la marche aléatoire, 64^2 points régulièrement espacés sur le carré unitaire ont été choisis pour calculer les sommes partielles des sauts générés.

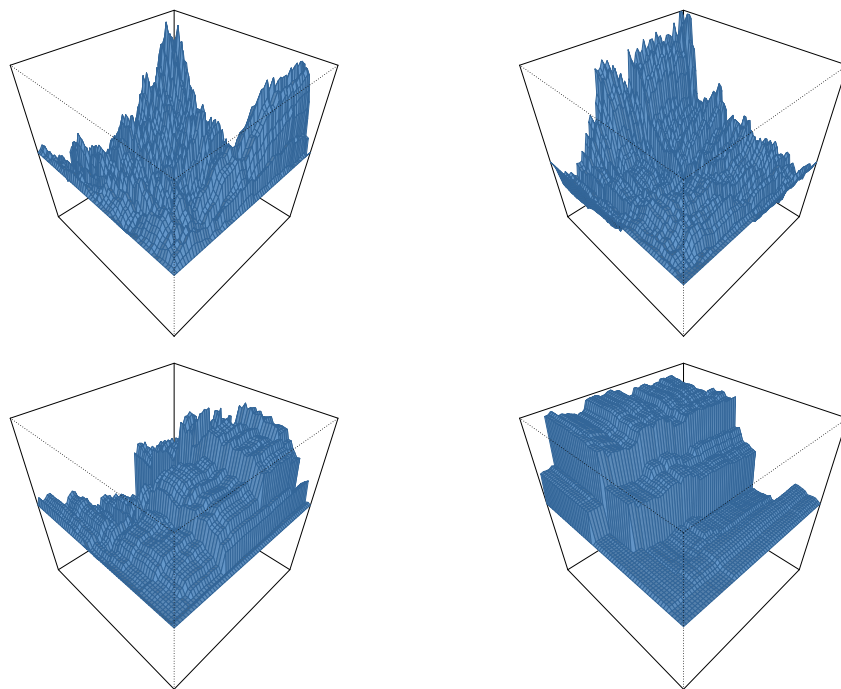


Figure 3.5. Différentes réalisations de draps stables. Le seul paramètre qui varie est l'indice de stabilité, α . En haut, à gauche $\alpha = 2$ et à droite $\alpha = 1.6$; en bas, à gauche $\alpha = 1.1$ et à droite $\alpha = 0.8$.

3.5 Ajustement d'un drap stable

Maintenant que le drap stable a été défini, la question qui vient naturellement à l'esprit est de savoir si l'ajustement des paramètres est possible, une fois que l'on a observé la réalisation d'un tel processus stochastique. La réponse est affirmative et la méthode se base sur la propriété suivante que l'on trouve dans Dalang and Khoshnevisan (2004) et qui découle de la construction théorique d'un tel drap.

Proposition 3.34 Soit $\{Z(s) : s \in \mathbb{R}_+^2\}$ un drap stable de paramètres α , β , σ et μ_0 . Pour tout $s, t \in \mathbb{R}_+^2$ tels que $s \leq t$, en posant $R = (s, t]$, on a que

$$\Delta_R Z \sim ((t_1 - s_1)(t_2 - s_2))^{1/\alpha} \cdot S_0(\alpha, \beta, \sigma, \tilde{\mu}_0), \quad (3.20)$$

où $\tilde{\mu}_0 = ((t_1 - s_1)(t_2 - s_2))^{1-1/\alpha} \mu_0 + \beta \sigma \tan\left(\frac{\pi\alpha}{2}\right) \left(1 - ((t_1 - s_1)(t_2 - s_2))^{1-1/\alpha}\right)$.

Dans cette section, la méthode mise au point pour ajuster les paramètres d'un drap stable basé sur une unique observation d'une réalisation est présentée. Afin de simplifier l'utilisation de la propriété présentée ci-dessus, on suppose que le drap est observé sur une grille régulière. La démarche représente le pendant bidimensionnel d'une étude classique d'un processus stochastique de Lévy unidimensionnel, où l'on considère des manipulations des accroissements reconstruits sur la base des observations en considérant des différences

d'observations successives. La marche à suivre se présente comme suit.

- I :** Extraction à partir de la réalisation observée d'un échantillon issu de variables aléatoires indépendantes et identiquement distribuées ;
- II :** Estimation par optimisation sous contrainte de la log-vraisemblance des paramètres de stabilité, d'asymétrie, de lieu et d'échelle.

3.5.1 Extraction d'un échantillon

Puisque l'on a supposé que le drap a été observé sur une grille régulière, cela signifie que l'on est en possession d'un échantillon qui peut s'écrire comme

$$\{\tilde{z}(s_i, s_j) : i = 0, 1, \dots, n, j = 0, 1, \dots, m; n, m \in \mathbb{N}\},$$

où les valeurs $\tilde{z}(s_i, s_j) := 0$ si $i = 0$ ou $j = 0$ ont été ajoutées artificiellement. Par conséquent, pour tout $i = 1, \dots, n$ et $j = 1, \dots, m$, on peut construire les incréments observés

$$\Delta_{i,j} = \tilde{z}(s_i, s_j) - \tilde{z}(s_{i-1}, s_j) - \tilde{z}(s_i, s_{j-1}) + \tilde{z}(s_{i-1}, s_{j-1}).$$

En posant $h_n = (s_i - s_{i-1})$, $i = 1, \dots, n$, $h_m = (s_j - s_{j-1})$, $j = 1, \dots, m$ et en s'aidant de (3.20), on sait que

$$\Delta_{i,j} \sim S_0(\alpha, \beta, \check{\sigma}, \check{\mu}_0),$$

où $\check{\sigma} = (h_n h_m)^{1/\alpha} \sigma$ et $\check{\mu}_0 = (h_n h_m) \mu_0 + \beta \sigma \tan\left(\frac{\pi\alpha}{2}\right) \left((h_n h_m)^{1/\alpha} - (h_n h_m)\right)$. Par conséquent, on a bien un échantillon, de taille égale à l'échantillon initial, composé d'observations sensées provenir de variables aléatoires indépendantes et identiquement distribuées selon une loi stable $S_0(\alpha, \beta, \check{\sigma}, \check{\mu}_0)$. En présence de cet échantillon, toute méthode d'estimation présentée à la Section 3.3.6 peut être appliquée afin d'ajuster les paramètres. Une fois les estimations obtenues pour $\check{\sigma}$ et $\check{\mu}_0$ et puisque les paramètres h_n et h_m sont connus, il est possible de récupérer les estimations de σ et de μ_0 en s'aidant de celles déjà obtenues. La méthode présentée ci-après est composée de différentes méthodes et s'inspire de celle mise au point par Mittnik *et al.* (1999).

3.5.2 Optimisation sous contrainte de la log-vraisemblance

La première idée était la suivante : estimer dans un premier temps les paramètres de lieu et d'échelle, standardiser l'échantillon et optimiser la fonction de log-vraisemblance associée aux paramètres restants, c'est-à-dire l'indice de stabilité et l'indice d'asymétrie sur la base d'un échantillon centré réduit. Comme l'ont révélé des simulations effectuées, cette

méthode bien qu'attrayante pour son gain de temps à comparer de la méthode basée sur la log-vraisemblance complète, se comporte bien pour autant que la distribution sous-jacente soit symétrique et/ou que l'indice de stabilité soit élevé. Dans les autres cas, la standardisation semble introduire un biais négatif, tant pour l'estimateur de α que pour celui concernant β . Cela résulte en partie du fait que, comme mentionné précédemment, l'estimation du paramètre d'échelle de la paramétrisation 0 d'une loi stable dépend également des estimations des paramètres α et β . En effet, lors des simulations, les valeurs initiales retenues pour ces deux paramètres étaient également utilisées pour obtenir une estimation du paramètre μ_0 .

Par conséquent, une optimisation de la fonction de log-vraisemblance complète s'est avérée devoir s'imposer. Si l'on note par $\theta = (\alpha, \beta, \sigma, \mu_0)^t$ le vecteur des paramètres dont on souhaite faire l'inférence, l'estimateur du maximum de vraisemblance de θ est obtenu en maximisant la fonction de log-vraisemblance

$$l(\theta; \Delta_1, \dots, \Delta_K) = \sum_{i=1}^K \log(f_0(\Delta_i; \theta)), \quad (3.21)$$

par rapport à θ , où $(\Delta_1, \dots, \Delta_K)$ est l'échantillon obtenu par la méthode présentée à la section précédente et $K = nm$. Dans cette écriture, la fonction $f_0(\cdot; \theta)$ correspond à la fonction de densité associée à la paramétrisation 0 d'une variable aléatoire stable de paramètres α , β , σ et μ_0 . DuMouchel (1973) a étudié les propriétés théoriques de l'estimateur du maximum de vraisemblance pour θ et a montré qu'il possédait une distribution asymptotique normale, pour autant que quelques conditions de régularités soient satisfaites (en fait, il est nécessaire de supposer que l'indice de stabilité vérifie $\alpha \geq \epsilon$ pour un $\epsilon > 0$). Autrement dit,

$$\sqrt{K}(\hat{\theta}_{\text{ML}} - \theta_o) \xrightarrow{D} \mathcal{N}(0, I^{-1}(\theta_o)), \quad (3.22)$$

où

$$I(\theta) = -E\left(\frac{\partial^2}{\partial\theta\partial\theta^t} l(\theta; \Delta_1, \dots, \Delta_K)\right) \quad (3.23)$$

représente la matrice d'information de Fisher. Dans ce travail comme dans Mittnik *et al.* (1999), la maximisation de (3.21) est effectuée de manière numérique. La première différence est la paramétrisation retenue. Contrairement à Mittnik *et al.* (1999) qui ont employé la paramétrisation 1, c'est la paramétrisation 0 qui a été retenue dans ce travail. Une des difficultés dans cette approche numérique vient du fait que l'optimisation est sensée s'effectuer sous contraintes. En effet, les conditions

$$0 < \alpha \leq 2, \quad -1 \leq \beta \leq 1, \quad \text{et} \quad \sigma > 0,$$

doivent être vérifiées. Plutôt que de résoudre cette optimisation sous contraintes, une reparamétrisation du problème permet de relâcher ces contraintes. En posant $\theta = h(\nu)$ avec $\nu = (\check{\alpha}, \check{\beta}, \check{\sigma}, \check{\mu}_0)^t$, cette transformation s'opère composante par composante et est donnée par

$$\begin{aligned}\alpha &= 2\exp\left(-(\check{\alpha}/a)^2\right), \\ \beta &= 2\exp\left(-(\check{\beta}/a)^2\right) - 1, \\ \sigma &= \exp(\check{\sigma}/a), \\ \mu_0 &= \check{\mu}_0,\end{aligned}\tag{3.24}$$

où $\check{\alpha}, \check{\beta}, \check{\sigma}$ et $\check{\mu}_0 \in \mathbb{R}$, et $a \in \mathbb{R}$. Il faut encore remarquer que même si en théorie $\alpha \in (0, 2]$, notre intérêt s'est porté sur des valeurs de ce paramètre supérieures à 1. Pour comparaison, la paramétrisation utilisée est représentée dans la Figure 3.6. Elle est comparée à la transformation choisie par Mittnik *et al.* (1999) dont le gradient est généralement beaucoup plus élevé.

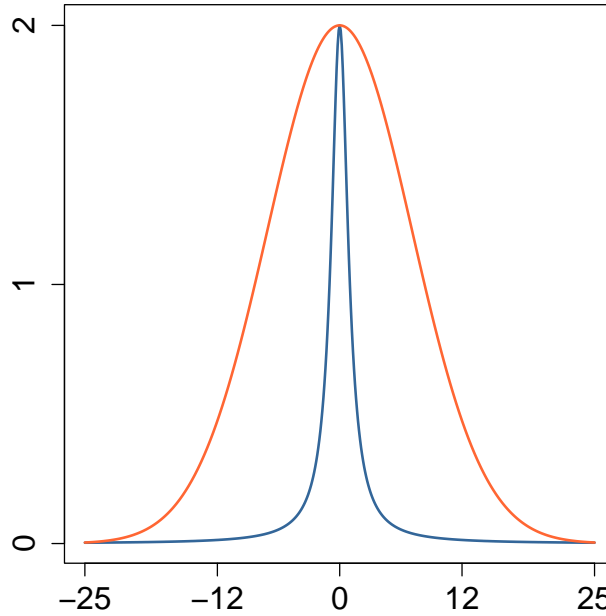


Figure 3.6. Comparaison des transformations permettant d'exprimer le paramètre α borné en fonction d'un paramètre $\check{\alpha}$ réel. En orange, la transformation appliquée dans ce travail ; en bleu celle utilisée par Mittnik *et al.* (1999).

Par conséquent l'équation (3.22) devient

$$\sqrt{K}(\hat{\theta}_{\text{ML}} - \theta_o) \xrightarrow{\text{D}} \mathcal{N}\left(0, \left(\frac{\partial}{\partial \nu^t} h(\nu)\right)\Big|_{\nu=\nu_o} I^{-1}(\nu_o) \left(\frac{\partial}{\partial \nu^t} h(\nu)\right)^t \Big|_{\nu=\nu_o}\right). \tag{3.25}$$

Dés lors, l'optimisation se décompose comme suit :

- Définition de valeurs initiales pour les paramètres ;
- Approximation de la densité par transformée de Fourier de la fonction caractéristique ;
- Application d'un algorithme de descente.

Les deux dernières étapes étant répétées jusqu'à ce qu'un critère d'arrêt soit satisfait.

Définition de valeurs initiales pour les paramètres

Pour les paramètres de stabilité et d'asymétrie, c'est une méthode basée sur celle proposée par McCulloch (1986) qui a été mise en place. Tout d'abord, les estimations associées aux estimateurs η_α et η_β donnés par (3.15) sont calculées sur la base de l'échantillon $(\Delta_1, \dots, \Delta_K)$. D'autre part, des simulations sont effectuées pour obtenir une estimation des valeurs prises par ces statistiques en fonction des paramètres α et β . Le résultat de ces simulations est présenté dans la Figure 3.7. Une fois l'estimation de η_α associée à l'échantillon $(\Delta_1, \dots, \Delta_K)$ calculée, on choisit pour valeur initiale $\hat{\alpha}_0$, la valeur de α qui minimise l'écart entre l'estimation obtenue et les droites de régression ajustées par simulation. Ensuite, sur la base de cette valeur initiale $\hat{\alpha}_0$, on calcule la valeur de η_β associée à l'aide des régressions ajustées par simulation. Une fois encore, la valeur $\hat{\beta}_0$ est celle qui minimise l'écart entre la valeur basée sur les données de la statistique η_β et les droites de régression ajustées.

Pour le paramètre d'échelle, σ , c'est l'estimateur proposé par Fama and Roll (1971) qui a été utilisé :

$$\hat{\sigma}_0 = \frac{\Delta_{(.72)} - \Delta_{(.18)}}{1.654}.$$

Cet estimateur possède un biais borné supérieurement par 4% si la distribution est symétrique ($\beta = 0$). Si tel n'est pas le cas, il est difficile de trouver une borne invariante sur le biais. Cet estimateur a toutefois été retenu, en raison de ses bonnes performances globales et du fait qu'il ne s'agit que d'obtenir une valeur initiale. Pour le paramètre de lieu, lorsque $1 < \alpha \leq 2$, la distribution stable possède une espérance finie. Qui plus est, dans la paramétrisation 1 d'une loi stable, le paramètre de lieu μ_1 correspond à l'espérance de la variable aléatoire. Par conséquent, la moyenne de l'échantillon est un estimateur consistant pour ce paramètre. Cependant, une estimation plus robuste est obtenue en substituant une moyenne tronquée à la moyenne classique. Dans ce travail, c'est une moyenne tronquée à 25% qui a été utilisée. Puisque c'est la paramétrisation 0 d'une loi stable qui a été retenue, ce n'est pas μ_1 que l'on souhaite estimer, mais bien μ_0 . La formule qui les lie est donnée par (3.12). Cependant, une fois de plus, comme il s'agit d'une valeur initiale seulement, $\hat{\mu}_0$ a été fixé à la moyenne tronquée qui présente, par conséquent un biais (si la loi n'est pas symétrique). Finalement, on obtient les valeurs initiales $\check{\alpha}_0$, $\check{\beta}_0$, $\check{\sigma}_0$ et $\check{\mu}_0$ en utilisant les relations fournies par (3.24).

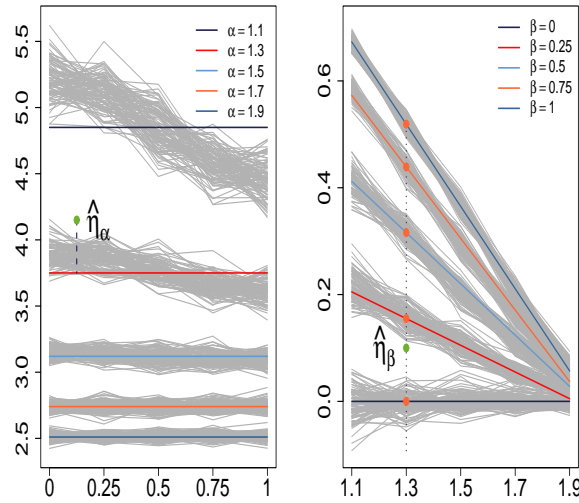


Figure 3.7. Simulation des estimateurs donnés par (3.15). A gauche, η_α , la statistique associée au paramètre α en fonction du paramètre β . Des simulations ont été effectuées pour différentes valeurs de α et β . Pour chaque valeur retenue de α , une droite de régression constante a été ajustée car on constate que la valeur de la statistique η_α ne dépend pas de la valeur prise par β , du moins lorsque α est grand. A droite, η_β , la statistique associée au paramètre β en fonction du paramètre α . Des simulations ont été effectuées pour différentes valeurs de α et β . Pour chaque valeur retenue de β , une droite de régression a été ajustée. Pour exemple, $\hat{\eta}_\alpha = 4.15$ (en vert à gauche) ; par conséquent, on choisit $\alpha_0 = 1.3$. D'autre part, $\hat{\eta}_\beta = 0.1$ (en vert à droite) ; par conséquent, on choisit $\beta_0 = 0.25$.

Approximation de la densité par transformée de Fourier

L'idée est d'écrire la densité d'une variable aléatoire stable comme la transformée de Fourier de sa fonction caractéristique. En effet, on a que

$$f_0(z; \alpha, \beta, \mu_0, \sigma) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-izu} \phi_0(u; \alpha, \beta, \mu_0, \sigma) du. \quad (3.26)$$

Par conséquent, on peut appliquer la méthode de la transformée de Fourier rapide pour obtenir la valeur de la densité associée à chaque observation de l'échantillon $(\Delta_1, \dots, \Delta_K)$. On procède comme suit. L'intégrale présentée dans (3.26) peut être calculée pour N points équi-répartis $\check{\Delta}_l = (l - 1 - N/2) \cdot h$, $l = 1, \dots, N$. En posant $u = 2\pi\omega$, on peut écrire

en s'aidant de (3.26) que

$$\begin{aligned} f_0(\check{\Delta}_l; \alpha, \beta, \mu_0, \sigma) &= f_0((l-1-N/2) \cdot h; \alpha, \beta, \mu_0, \sigma) \\ &= \int_{-\infty}^{\infty} e^{-i2\pi\omega(l-1-N/2)h} \phi_0(2\pi\omega; \alpha, \beta, \mu_0, \sigma) d\omega. \end{aligned}$$

En discrétisant, on peut approcher cette dernière intégrale par

$$f_0((l-1-N/2) \cdot h) \approx v \sum_{n=1}^N e^{-i2\pi(n-1-N/2)(l-1-N/2)hv} \phi_0(2\pi v(n-1-N/2)).$$

En posant $v = (hN)^{-1}$, on peut encore écrire

$$f_0(\check{\Delta}_l; \alpha, \beta, \mu_0, \sigma) \approx v(-1)^{l-1-N/2} \sum_{n=1}^N (-1)^{n-1} e^{-i2\pi(n-1-N/2)} \phi_0(2\pi v(n-1-N/2)). \quad (3.27)$$

Dans l'Equation (3.27), la somme présente dans le terme de droite peut être calculée de manière efficiente à l'aide de la méthode de la transformée de Fourier rapide appliquée à la séquence de valeurs

$$(-1)^{n-1} \phi_0(2\pi v(n-1-N/2)), \quad n = 1, \dots, N.$$

En normalisant le $l^{\text{ème}}$ élément de la séquence obtenue par $v(-1)^{l-1-N/2}$, on obtient une estimation de la densité au point $\check{\Delta}_l$. Finalement, pour obtenir la valeur de la densité aux points de l'échantillon $(\Delta_1, \dots, \Delta_K)$, une interpolation de degré 1 est effectuée sur la base des valeurs de la densité associées aux points équis-répartis. Les paramètres N et h peuvent varier en fonction de la précision souhaitée. Les valeurs $N = 2^{13}$ et $h = 0.01$ semblent cependant représenter un bon compromis entre précision et rapidité.

Application d'un algorithme de descente

Une fois que l'on possède la valeur de la densité pour tout point Δ_k , $k = 1, \dots, K$, on peut aisément évaluer la fonction de log-vraisemblance $l(\theta; x_1, \dots, x_K)$ exprimée dans (3.21) et par conséquent, évaluer la fonction de log-vraisemblance $\check{l}(\nu; \Delta_1, \dots, \Delta_K)$ associée à ν . Pour optimiser cette fonction, la méthode de descente, dite de Newton, a été utilisée. A chaque pas, elle nécessite le calcul du gradient et de la matrice hessienne associés à la fonction. En effet, si ν_p est la valeur du paramètre d'intérêt au $p^{\text{ème}}$ pas, alors ν_{p+1} , la valeur du paramètre d'intérêt au $(p+1)^{\text{ème}}$ pas est donné par

$$\nu_{p+1} := \nu_p - \check{H}(\nu_p)^{-1} \nabla \check{l}(\nu_p; \Delta_1, \dots, \Delta_K), \quad p \in \mathbb{N}, \quad (3.28)$$

où $\nabla \tilde{l}(\nu; \Delta_1, \dots, \Delta_K)$ représente le gradient de la fonction de log-vraisemblance associée à ν et $\tilde{H}(\nu)$ sa matrice hessienne. Cette méthode de descente est poursuivie jusqu'à ce qu'un critère d'optimalité soit rempli; par exemple, que la norme du vecteur des innovations soit suffisamment faible. Puisque la densité de variables aléatoires stables n'existe pas sous forme explicite, il est impossible d'obtenir une écriture explicite du gradient de la fonction de log-vraisemblance ou de la hessienne, à plus forte raison après un changement de paramétrisation. Ces valeurs doivent donc, par conséquent, être approchées. Pour ce faire, on s'aide des formules données ci-dessous.

$$\nabla \tilde{l}(\nu; \Delta_1, \dots, \Delta_K) = \left(\frac{\partial}{\partial \nu^t} h(\nu) \right) \cdot \nabla l(\theta; \Delta_1, \dots, \Delta_K), \quad (3.29)$$

$$\tilde{H}(\nu) = \left(\frac{\partial}{\partial \nu^t} h(\nu) \right) \cdot H(\theta) \cdot \left(\frac{\partial}{\partial \nu^t} h(\nu) \right)^t + A(\nu), \quad (3.30)$$

où $\nabla l(\theta; \Delta_1, \dots, \Delta_K)$ représente le gradient de la fonction de log-vraisemblance associée à θ , $H(\theta)$ sa hessienne et la matrice A est donnée par

$$A(\nu) = \left(\sum_{k=1}^4 \frac{\partial^2}{\partial \nu_i \partial \nu_j^t} \theta_k \cdot \frac{\partial}{\partial \theta_k} l(\theta; \Delta_1, \dots, \Delta_K) \right)_{1 \leq i, j \leq 4}.$$

Dans ces dernières écritures, les composantes du gradient $\nabla l(\theta; \Delta_1, \dots, \Delta_K)$ et celles de la hessienne $H(\theta)$ sont approchées par différences finies. Autrement dit, on approche les termes du gradient comme

$$\nabla l(\theta; \Delta_1, \dots, \Delta_K)_j \approx \frac{1}{r} \left(l((\dots, \theta_j - r/2, \dots)) - l((\dots, \theta_j + r/2, \dots)) \right), \quad (3.31)$$

et

$$\begin{aligned} H(\theta)_{ij} \approx \frac{1}{r^2} & \left(l((\dots, \theta_i - r/2, \dots, \theta_j - r/2, \dots)) + l((\dots, \theta_i + r/2, \dots, \theta_j + r/2, \dots)) \right. \\ & \left. - l((\dots, \theta_i + r/2, \dots, \theta_j - r/2, \dots)) - l((\dots, \theta_i - r/2, \dots, \theta_j + r/2, \dots)) \right), \end{aligned} \quad (3.32)$$

pour tout $1 \leq i, j \leq 4$. La valeur $r = 0.1$ a été utilisée pour les simulations et l'application. Cette approximation quadratique de la fonction de log-vraisemblance renforce l'importance de pouvoir évaluer la valeur de la fonction de log-vraisemblance de manière rapide. En effet, pour pouvoir approcher à l'aide de (3.31) et (3.32) l'ensemble des composantes du gradient et de la hessienne associés à un quadruplet θ , la procédure nécessite l'évaluation de log-vraisemblance pour pas moins de quarante autres quadruplets distincts. C'est d'ailleurs certainement une des raisons pour laquelle ce genre de méthode n'a vu le jour que tardivement.

Dans la section suivante, des simulations effectuées pour juger du comportement de l'estimateur proposé sont présentées.

3.6 Simulations

Lors de ces simulations, l'estimateur présenté dans la section précédente est comparé à l'estimateur mis en place par Press (1972), présenté dans la Section 3.3.6. Le nombre de répétitions a été fixé à $r = 2'000$. Chaque simulation consiste dans un premier temps à générer un drap stable stationnaire sur le carré unitaire. La suite de variables aléatoires indépendantes et identiquement distribuées générées sur le carré est composée d'un million de termes. Pour la marche aléatoire, les paramètres n et m ont été tous les deux fixés à 32. Autrement dit, le drap stable est observée à 1024 sites. Une fois le drap généré, l'échantillon basé sur les incréments planaires est calculé à l'aide de la formule donnée par (3.20). La taille de l'échantillon est par conséquent égale au nombre de sites où le drap a été observé. Les valeurs des paramètres varient comme suit

$$\left\{ \begin{array}{l} \alpha \in \{1.2, 1.4, 1.6, 1.8\}, \\ \beta \in \{0, 0.3, 0.6\}, \\ \mu_0 \in \{0, 3\}, \\ \sigma \in \{1, 4\}. \end{array} \right.$$

Pour chaque combinaison de valeurs des paramètres, 500 répétitions ont été effectuées. La Figure 3.8 compare les diagrammes en boîtes obtenus pour les deux estimateurs pour chacun des paramètres dans le cas où $\mu_0 = 0$ et $\sigma = 1$. Chaque graphique correspond à une valeur fixée du paramètre d'asymétrie β et contient, de gauche à droite, les estimations obtenues lorsque le paramètre de stabilité α prend les différentes valeurs possibles. On constate que pour chacun des paramètres à estimer, les deux estimateurs se comportent globalement bien. Concernant le paramètre de stabilité, l'estimateur de Press semble sans biais mais possède de toute évidence une plus grande variance que l'estimateur du maximum de vraisemblance. Ce dernier, par contre, semble admettre un léger biais positif, spécialement lorsque la distribution est fortement asymétrique. Pour le paramètre d'asymétrie, l'estimateur du maximum de vraisemblance semble d'autant mieux estimer la valeur de β que le paramètre de stabilité est faible. Cela traduit le fait que le paramètre d'asymétrie perd de son importance à mesure que α se rapproche de 2. Pour l'estimateur de Press, c'est le contraire. Concernant le paramètre de lieu, les deux estimateurs semblent se comporter de manière similaire. L'avantage revient tout de même à l'estimateur du maximum de vraisemblance qui exhibe une variabilité plus faible. Finalement, concernant le paramètre d'échelle, l'estimateur de Press se déprécie en même temps que

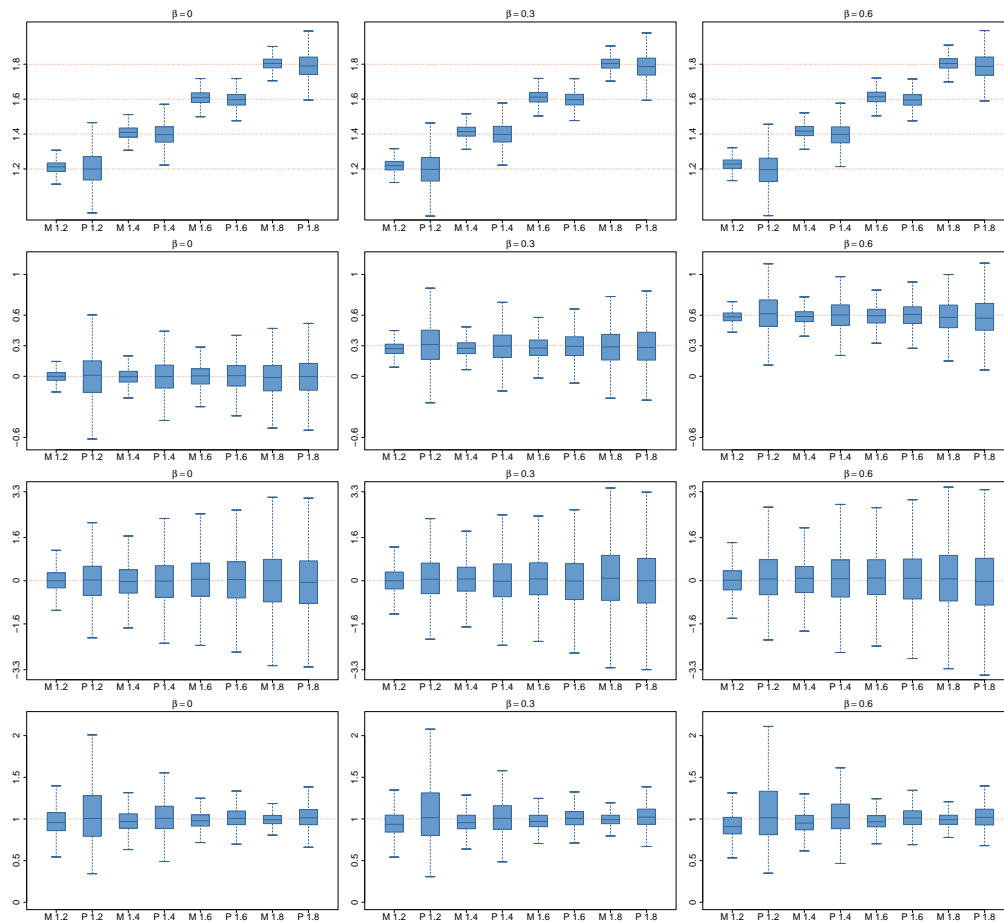


Figure 3.8. Cas $\mu_0 = 0$ et $\sigma = 1$: Diagrammes en boîte pour les différents paramètres. La première ligne correspond au paramètre α ; la seconde, au paramètre β ; la troisième, au paramètre μ_0 et la dernière au paramètre σ . La première colonne concerne les cas où $\beta = 0$; la deuxième, les cas où $\beta = 0.3$ et la troisième, les cas où $\beta = 0.6$. Dans chaque graphique, de gauche à droite, le paramètre α varie entre 1.2 et 1.8. Les diagrammes en boîte sont représentés par paire ; à gauche l'estimateur du maximum de vraisemblance, à droite, celui de Press.

le paramètre de stabilité diminue. Il en va de même pour l'estimateur du maximum de vraisemblance mais la péjoration est plus faible.

Le Tableau 3.1 permet de comparer les carrés moyens associés aux deux estimateurs pour chacune des combinaisons des paramètres α et β lorsque $\mu_0 = 0$ et $\sigma = 1$. On constate que, relativement à ce critère, l'estimateur basé sur la vraisemblance est bien meilleur que l'estimateur de Press (1972). La Figure 3.9 et le Tableau 3.2 illustrent le cas où $\mu = 3$ et $\sigma = 4$. Le comportement des deux estimateurs semble relativement similaire en ce qui concerne l'estimation des paramètres de stabilité et d'asymétrie. La variabilité lors de l'estimation des paramètres d'échelle et de lieu, par contre, augmente considérablement. Si on compare les Tableaux 3.1 et 3.2, on constate que les carrés moyens associés aux paramètres α et β pour l'estimateur du maximum de vraisemblance sont

		$\alpha = 1.2$		$\alpha = 1.4$		$\alpha = 1.6$		$\alpha = 1.8$	
		ML	PR	ML	PR	ML	PR	ML	PR
$\beta = 0.0$	CME_α	2 (3)	10 (14)	2 (3)	5 (7)	3 (4)	3 (4)	2 (5)	6 (11)
	CME_β	3 (5)	49 (65)	6 (9)	28 (38)	12 (16)	20 (27)	35 (51)	39 (59)
	CME_{μ_0}	165 (235)	629 (902)	416 (558)	750 (1072)	758 (1036)	925 (1261)	1324 (1774)	1363 (1851)
	CME_σ	36 (50)	140 (249)	23 (31)	53 (90)	15 (22)	20 (34)	8 (13)	22 (38)
$\beta = 0.3$	CME_α	2 (4)	9 (13)	2 (4)	5 (7)	2 (3)	3 (4)	2 (5)	6 (10)
	CME_β	5 (7)	43 (58)	7 (9)	25 (34)	12 (16)	19 (26)	32 (46)	42 (66)
	CME_{μ_0}	198 (273)	680 (1003)	410 (565)	761 (1064)	795 (1125)	993 (1421)	1451 (2027)	1468 (2047)
	CME_σ	36 (45)	143 (247)	23 (32)	52 (87)	14 (21)	18 (29)	8 (13)	22 (38)
$\beta = 0.6$	CME_α	3 (4)	9 (13)	3 (4)	5 (7)	2 (3)	3 (4)	2 (5)	5 (7)
	CME_β	3 (5)	34 (44)	5 (8)	21 (28)	10 (14)	15 (21)	25 (34)	46 (76)
	CME_{μ_0}	282 (376)	906 (1419)	525 (721)	963 (1285)	959 (1330)	1181 (1653)	1485 (1983)	1529 (2043)
	CME_σ	36 (41)	145 (242)	23 (30)	51 (83)	14 (19)	19 (32)	7 (12)	21 (36)

Tableau 3.1. Carré moyen de l'erreur ($\times 10^3$) pour le cas où $\mu_0 = 0$ et $\sigma = 1$ en fonction des valeurs de α et β . En-dessous, entre parenthèses, l'écart-type associé. Nombre de répétitions : $r = 2000$.

semblables, tandis que pour l'estimateur de Press, le carré moyen de l'erreur associé à β a également augmenté. Dans la section suivante, ces deux estimateurs sont utilisés pour une application pratique sur des données agricoles.

3.7 Application

L'objectif de cette section est de démontrer l'intérêt des draps stables lors d'une application. La méthode d'estimation présentée dans la Section 3.5 et la nature même des draps de Lévy a forcé la considération d'un jeu de données dont les observations ont été obtenues sur une grille régulière. Les données retenues pour cette application sont celles fournies par Mercer and Hall (1911), également présentées et traitées par Cresie (1993, p.454). Ce jeu de données concerne le rendement d'un champ de blé. Ce dernier, d'une superficie total d'un acre (0.4 hectare), a été divisé en cinq cents parcelles de superficie égale. Le blé produit et récolté sur chaque parcelle a été pesé et les données représentent le rendement en livre de chacune des parcelles. En définissant le

		$\alpha = 1.2$		$\alpha = 1.4$		$\alpha = 1.6$		$\alpha = 1.8$	
		ML	PR	ML	PR	ML	PR	ML	PR
$\beta = 0.0$	CME_α	3 (4)	27 (39)	2 (3)	15 (21)	2 (3)	7 (10)	2 (4)	3 (4)
	CME_β	3 (5)	150 (191)	6 (8)	100 (129)	11 (15)	72 (95)	34 (48)	72 (95)
	CME_{μ_0}	2958 (4132)	15358 (23533)	7039 (9593)	17834 (25895)	13277 (18512)	20159 (28241)	22669 (31267)	24537 (33806)
	CME_σ	601 (861)	3882 (6812)	392 (585)	1739 (3123)	240 (364)	640 (1119)	135 (221)	196 (319)
$\beta = 0.3$	CME_α	3 (4)	27 (39)	2 (4)	14 (20)	2 (3)	8 (11)	2 (4)	3 (4)
	CME_β	5 (7)	139 (180)	8 (11)	84 (109)	12 (16)	65 (89)	31 (46)	63 (84)
	CME_{μ_0}	3768 (5433)	16378 (25535)	7551 (10266)	17362 (24583)	13895 (19106)	20080 (29096)	23711 (34262)	23678 (32752)
	CME_σ	548 (710)	4154 (7822)	366 (542)	1694 (3109)	249 (384)	701 (1224)	130 (201)	194 (314)
$\beta = 0.6$	CME_α	4 (5)	27 (41)	3 (4)	15 (21)	2 (3)	8 (10)	2 (5)	3 (4)
	CME_β	4 (5)	108 (138)	6 (8)	64 (77)	10 (13)	48 (57)	25 (34)	47 (64)
	CME_{μ_0}	6696 (10499)	18625 (29898)	9313 (13367)	19608 (29764)	15365 (21041)	21454 (30552)	23257 (31238)	23544 (31307)
	CME_σ	546 (679)	3932 (7582)	361 (445)	1664 (2825)	219 (290)	682 (1138)	129 (208)	184 (273)

Tableau 3.2. Carré moyen de l'erreur ($\times 10^3$) pour le cas où $\mu_0 = 3$ et $\sigma = 4$ en fonction des valeurs de α et β . En-dessous, entre parenthèses, l'écart-type associé. Nombre de répétitions : $r = 2000$.

sommet sud ouest de ce champ pour origine, on peut sommer ces rendements en les considérant comme des incréments spatiaux et ainsi obtenir un processus stochastique bidimensionnel comme présenté dans la Figure 3.10, à gauche. Il serait injustifié de parler à ce stade de drap, puisque une dépendance spatiale existe. Ce jeu de données a été traité par différentes méthodologies. Dans le domaine spatial, Cressie (1993) propose une modélisation géostatistique ainsi qu'une modélisation par treillis. Puisque c'est le thème même de ce travail, une modélisation géostatistique des incréments a été retenue. Ces incréments qui correspondent aux données fournies par Mercer and Hall (1911), sont illustrées dans la Figure 3.12, en haut à gauche.

La tendance spatiale est obtenue à l'aide d'un lissage par médiane. La surface obtenue est présentée dans la Figure 3.12. L'effet associé aux colonnes est nettement plus prononcé que celui associé aux lignes de la grille d'observation. Une fois cette tendance obtenue, elle est soustraite aux observations. A partir de ces nouvelles observations, un processus intrinséquement stationnaire est ajusté par krigeage ordinaire. Pour ce faire, la dépendance spatiale isotropique est estimée de manière robuste et un variogramme exponentiel est

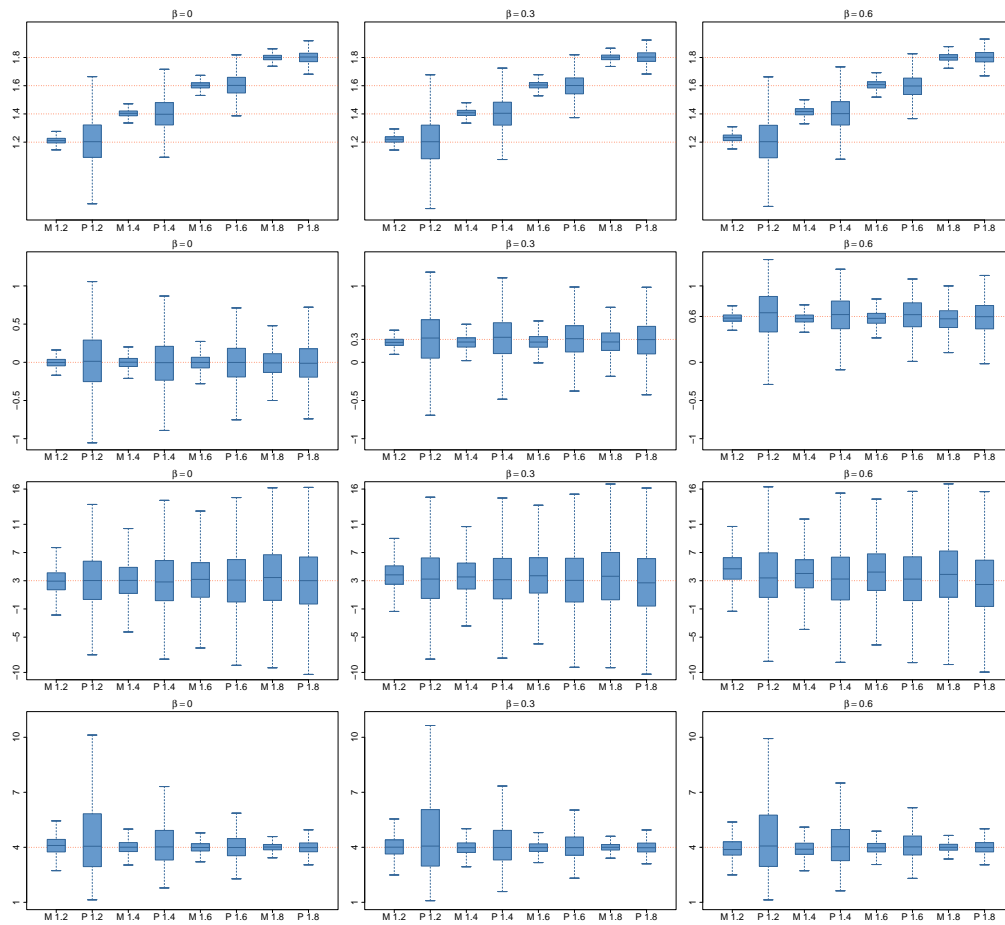


Figure 3.9. Cas $\mu_0 = 3$ et $\sigma = 4$: Diagrammes identique à la Figure 3.8.

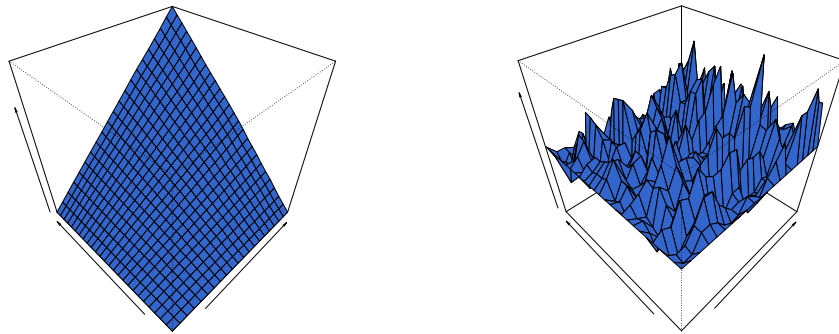


Figure 3.10. A gauche, le processus stochastique bidimensionnel obtenu en sommant les données fournies par Mercer and Hall (1911). Assez logiquement, le rendement augmente avec la superficie considérée. A droite, le drap de Lévy obtenu en agrégeant les résidus résultant d'un ajustement géostatistique des données. Ces résidus en questions sont présentés dans la Figure 3.12, en bas à droite.

ajusté. Une représentation du variogramme obtenue est fournie par la Figure 3.11. La valeur relativement élevée de l'effet-noyau est probablement représentative de l'imprécision sur la localisation des données puisque ce sont des données agrégées par parcelle.

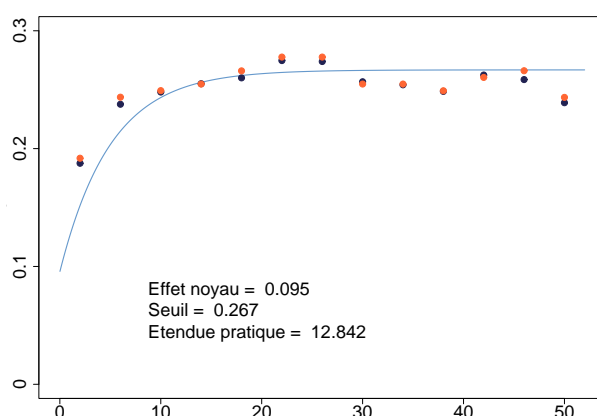


Figure 3.11. Ajustement du variogramme pour les données de Mercer and Hall (1911). Les deux estimateurs considérés sont ceux fournis par les Equations (1.11) et (1.12). On constate que dans ce cas, les estimations obtenues sont passablement similaires. L'ajustement d'un variogramme exponentiel s'est fait par moindres carrés pondérés.

Une fois le variogramme ajusté, une prédiction spatiale est effectuée par krigeage ordinaire. Lorsque la localisation de la prédiction coïncide avec un site d'observation, une perturbation infinitésimale a été introduite afin d'éviter une interpolation exacte, propre au krigeage ordinaire. La surface obtenue en sommant la tendance et la surface prédite par krigeage ordinaire est représentée dans la Figure 3.12, en bas à gauche. Si on soustrait aux données originales ces valeurs prédites, on obtient les résidus présentés dans la Figure 3.12, en bas à droite. Si l'on considère ces résidus comme les incréments indépendants d'un drap de Lévy, on obtient la surface représentée dans la Figure 3.10, à droite.

Considérer que les résidus obtenus constituent les incréments d'un drap de Lévy se justifie par la nature des observations et, par conséquent, la leur. En effet, ces données représentent un rendement par parcelle; autrement dit, elles sont de nature agrégée. Comme le mentionne Cressie (1993, p.248) dans sa présentation des données, ce détail peut poser une légère difficulté quant à l'utilisation de méthode géostatistique. Par contre, ce même détail vient justifier l'ajustement d'un drap stochastique. La dépendance spatiale ayant été supprimée, les incréments de ce drap (les résidus) peuvent être considérés comme les réalisations de variables aléatoires deux à deux indépendantes. A l'aide de la méthode présentée dans la Section 3.5, un ajustement d'un drap stable à la surface présentée dans la Figure 3.10, à droite, est effectuée. Les estimations obtenues sont données dans le Tableau 3.3.

Afin de tester si l'estimation obtenue pour le paramètre de stabilité est significativement différente de la valeur 2, ou autrement dit, si le drap ajusté est significativement éloigné d'un drap brownien, des intervalles de confiance ont également été calculés et sont présentés dans le Tableau 3.4. Ils représentent les intervalles de confiance à 95% pour les

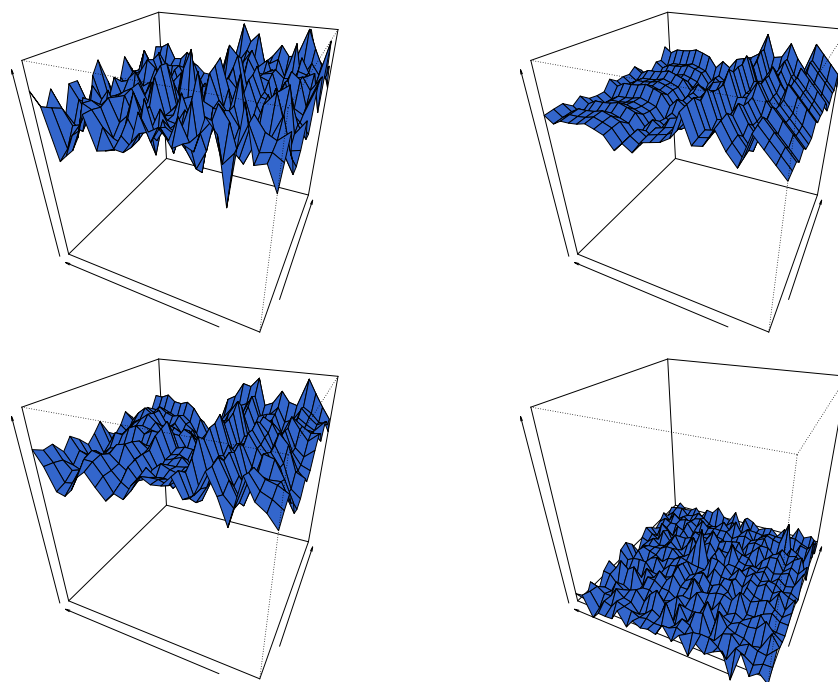


Figure 3.12. En haut, à gauche, une représentation des données considérées comme les incréments spatiaux de la surface présentée dans la Figure 3.10, à gauche. En haut à droite, la tendance obtenue à l'aide d'un lissage par médiane. En bas, à gauche, la surface ajustée composée de la tendance et de la surface obtenue par krigeage ordinaire appliqué aux données sans tendance. En bas, à droite, les résidus, considérés comme les incréments du drap de Lévy présenté dans la Figure 3.10, à droite.

	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\mu}_0$
MVE	1.895	-0.078	0.139	-0.002
PRESS	1.901	-0.070	0.139	-0.002

Tableau 3.3. Estimations des paramètres du drap stable associés aux résidus du traitement géostatistique des données de Mercer and Hall (1911) obtenues par la méthode mise au point et présentée à la Section 3.5 et par la méthode de Press (1972).

paramètres basés sur les résidus de l'ajustement géostatistique des données de Mercer and Hall (1911).

On constate plusieurs choses. Tout d'abord que les estimations de moyenne et de variance obtenues par application de la méthode des moments aux résidus sont bels et bien contenus dans les intervalles associés à μ_1 et σ . Pour constater cela, il est utile d'employer la transformation donnée par (3.12) et de se souvenir qu'entre le paramètre σ d'une loi stable $S_0(2, 0, \sigma, \mu)$ et l'écart-type de la loi normale associée, il y a un facteur multiplica-

	Borne inf.	Borne sup.
α	1.800	1.990
β	-1	1
σ	0.112	0.167
μ_0	-0.004	-0.001

Tableau 3.4. Intervalles de confiance à 95% pour les paramètres du drap ajusté sur la base des résidus de prédiction.

tif de $1/\sqrt{2}$. Plus intéressant, le fait que l'intervalle associé au paramètre d'asymétrie β couvre l'ensemble des valeurs possibles renforce le fait que ce paramètre perd de sa signification à mesure que le paramètre de stabilité se rapproche du cas gaussien. Même pour un niveau plus élevé, l'intervalle de confiance associé à β dépasse largement les limites de son domaine. Finalement, on constate que la valeur 2 est bien en dehors de l'intervalle de confiance associé au paramètre α . On rejette par conséquent que le drap considéré est un drap brownien.

Une autre manière de tester si l'indice de stabilité est significativement différent de la valeur 2 associée au cas gaussien est de considérer la fonction de vraisemblance profilée pour α et de baser la construction d'un intervalle de confiance pour le paramètre d'intérêt sur la statistique du rapport de vraisemblance. Les résultats de cette méthode sont présentés dans la Figure 3.13. On constate que l'intervalle de confiance pour α trouvé, $[1.782, 1.984]$, ne couvre à nouveau pas la valeur $\alpha = 2$.

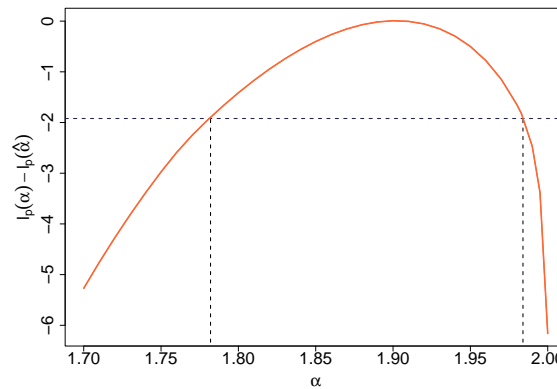


Figure 3.13. Logarithme du rapport de vraisemblance profilée pour α . La ligne horizontale correspond à la valeur $-\chi_1^2(0.95)/2$. Par conséquent les segments verticaux fournissent un intervalle à 95% pour le paramètre α : $[1.782, 1.984]$.

Une dernière illustration de ce fait est donné par la Figure 3.14. A gauche, les erreurs sont standardisées à l'aide de la variance de krigeage fournie par (1.3). A droite, les erreurs sont standardisées à l'aide de la formule donnée par (3.13). Les traits bleus

représentent dans les deux cas les droites d'équation $y = x$. Leurs extrémités sont données par les quantiles à 0.1% et leurs symétriques d'une loi normale standard et d'une loi stable standard. Pour le cas stable, on constate que la droite s'ajuste bien aux résidus standardisés, même les plus éloignés, qu'elle couvre l'ensemble de l'échantillon et qu'elle est même conservatrice à gauche. Pour le cas normal, il est un peu surprenant que les résidus de prédiction standardisés ne soient pas du tout alignés avec le trait bleu qui représente la droite d'équation $y = x$. Ceci peut en partie s'expliquer par le fait que la variance de krigeage donnée par (1.3), malgré son appellation, ne représente pas vraiment une estimation de la variance de prédiction, mais plus un indice de configuration des données comme le souligne Srivastava (1986).

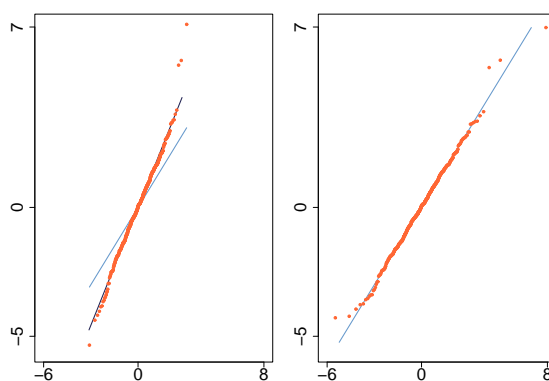


Figure 3.14. Diagrammes des quantiles associés aux résidus. À gauche, le cas normal ; à droite, le cas stable. À gauche, les erreurs sont standardisées à l'aide de la variance de krigeage fournie par (1.3). À droite, les erreurs sont standardisées à l'aide de la formule donnée par (3.13). Les traits bleus représentent dans les deux cas les droites d'équation $y = x$. Leurs extrémités sont données par les quantiles à 0.1% et leurs symétriques d'une loi normale standard et d'une loi stable standard. Dans le graphique de gauche, le trait noir correspond à un ajustement d'une régression linéaire des résidus standardisés en fonction des quantiles d'une loi normale standard. Ses extrémités sont données par le quantile à 10^{-6} et son symétrique d'une loi normale standard.

Mis à part ce problème d'échelle, les résidus de prédiction standardisés s'alignent tout de même relativement bien. En témoigne le trait noir représentant un ajustement d'une régression linéaire des résidus de prédiction standardisés en fonction des quantiles d'une loi normale standard. Ses extrémités sont données par le quantile à 10^{-6} et son symétrique d'une loi normale standard. Et c'est là le problème. Non seulement, le trait noir ne couvre pas l'entier de l'échantillon, mais, en plus, les erreurs standardisées se trouvent en-dessous du trait à gauche et au-dessus à droite, signifiant que les queues de la distribution empirique des erreurs standardisées sont plus lourdes que celles d'une loi normale. Cressie (1993, p.255) qui obtient étrangement un meilleur ajustement affirme

qu'il convient d'éditer les trois valeurs qui correspondent aux résidus de prédiction standardisés les plus éloignés à droite car elle présente des valeurs anormales relativement à leurs voisines. Il se propose de les éditer de sorte que les résidus de prédiction standardisés associés se retrouvent après édition sur la droite d'équation $y = x$. Ceci n'est pas forcément souhaitable et ne suffit pas à répondre à toutes les imperfections de cette modélisation gaussienne des résidus. L'édition *a posteriori* des données est un exercice périlleux, spécialement si on entend effectuer de l'inférence par la suite. La modélisation par un drap stable semble donc, dans cet exemple, tout à fait à propos.

Conclusion

Les lignes qui suivent ont une saveur toute particulière puisqu'elles sont sensées mettre un terme à ce travail et, par conséquent, à une période de ma vie. En les écrivant, les sentiments de l'accomplissement et de la frustration se mélangent, heureusement de manière inéquitable. Tout d'abord, la satisfaction ; celle d'être passé au travers d'une expérience riche à de nombreux niveaux et surtout intellectuellement stimulante. Satisfaction encore, en pensant aux nouveaux défis que je vais pouvoir relever, en rapport ou non avec ce travail. Une légère amertume ensuite ; celle liée à l'idée de clore ces recherches au moment où ma perception des phénomènes étudiés est la plus émoussée. Je souhaite ci-dessous revenir sur les différentes parties de ce travail et synthétiser le travail effectué. Je souhaite également énoncer les pistes qu'il m'aurait apparu sensé de poursuivre à partir du point atteint.

Dans le Chapitre 1, j'ai démontré que je possédais une bonne connaissance de la méthodologie classique de la géostatistique. Cela peut paraître réducteur, tant il est vrai qu'une présentation exhaustive nécessiterait un ouvrage complet. Cependant, il faut plutôt y voir un état de l'art qui s'est avéré nécessaire pour entreprendre la suite de ce travail.

Le Chapitre 2 peut se diviser en trois parties. La première dans laquelle une nouvelle modélisation est proposée. Celle-ci est motivée par la présence plus fréquente que rare de contamination dans les jeux de données. Alors que la modélisation à erreurs additives apparaît plus comme alibi nécessaire à l'élaboration de nouveaux outils géostatistiques, la modélisation à erreurs substitutives traduit plus clairement la perception qu'ont les scientifiques d'une contamination et permet d'y répondre de manière adéquate. Cette proposition fait partie des innovations de ce travail. La deuxième partie correspond à la présentation de l'estimateur de krigeage théorique associé au modèle. Par krigeage, il est entendu estimateur optimal au sens des moindres carrés. Ses propriétés de base ont été énoncées dans ce chapitre. Il exhibe une grande flexibilité et peut tout à fait être appliqué à des jeux de données propres. Dans ce cas, il se confond avec l'estimateur du

krigeage ordinaire. La dernière partie de ce chapitre correspond à la mise sur pied d'une approximation de l'estimateur, complètement automatisée et calculable en un temps raisonnable. Cet estimateur comble un vide dans le sens où il permet de traiter des données dépendantes qui présentent des valeurs aberrantes. Si une élimination sélective a priori peut encore se justifier pour un échantillon indépendant, elle est beaucoup plus critiquable en présence de dépendance, qui plus est spatiale. L'automatisation permet également des applications dans la surveillance de phénomènes naturels comme l'illustre l'application. Des simulations ont également été présentées pour étayer ses bonnes performances. Mise à part l'optimisation du code proposé (dans R), je pense qu'un sujet intéressant pour poursuivre dans le sujet serait l'élaboration d'un test formel pour la séparation des données en deux classes, les contaminées et les autres. Ceci n'est probablement pas simple, puisqu'un tel test devrait se baser sur la distribution d'extrêmes d'un échantillon de statistiques dépendantes. Qui plus est, la caractérisation de cette dépendance n'apparaît pas comme triviale. Si au départ, l'idée de cette nouvelle modélisation peut paraître farfelue ou de l'ordre du détail, la méthodologie qui en découle et qui a été proposée tout au long de ce chapitre devrait, je pense, permettre de reconsidérer ce jugement. Son intérêt, dans le milieu académique peut malheureusement être dévalorisée en raison des difficultés à dériver de manière théorique les caractéristiques de l'estimateur associé. Ce point peut à mon avis aisément être contrebalancé par l'engouement que pourrait rencontrer la méthodologie dans les applications concrètes.

Le Chapitre 3 constitue la seconde partie de mon travail. Elle a été motivée par la nécessité de proposer une modélisation qui permette de rendre l'hétérogénéité que peut présenter une surface. Un exemple d'une telle surface est donnée par celle constituée des résidus obtenus après un ajustement géostatistique robuste, comme celui présenté dans le chapitre précédent. Ce chapitre peut à son tour être divisé. Dans un premier temps, j'ai précisé l'origine des distributions stables et brossé un panorama relativement exhaustif du traitement statistique de telles distributions. J'ai également présenté la définition d'un drap stable. Dans un second temps, j'ai mis sur pied une méthodologie d'estimation des paramètres d'un tel drap. Se basant sur les propriétés du drap, la méthode élaborée est innovatrice de par la paramétrisation retenue et la programmation de toute la partie optimisation. Ceci a été effectué dans le langage C et peut être chargé dans le logiciel R. La méthode d'estimation ne connaît, à ma connaissance, pas de concurrence parmi les outils à disposition dans ce logiciel; le seul concurrent envisageable effectuant de l'intégration directe sans se soucier des contraintes temporelles. Autrement dit, la méthode proposée ne serait pas envisageable avec les seuls outils à disposition. Des simulations comparent l'estimateur proposé à un autre estimateur qui lui est beaucoup plus direct au niveau des calculs. Les résultats semblent toutefois indiquer que l'utilisation d'un temps de calcul supplémentaire raisonnable peut tout de même s'avérer intéressant en terme

de la qualité de l'estimation obtenue. Finalement, une application est présentée. Pour conclure les remarques concernant ce chapitre, il faut faire preuve d'honnêteté et accorder le fait qu'un drap stable ne s'applique pas aussi facilement qu'espéré. Ses deux principaux défauts sont les suivants. Tout d'abord, l'importance attribuée théoriquement aux deux axes de la base canonique de \mathbb{R}^2 est une caractéristique qui se retrouve difficilement dans la nature. Deuxièmement, le fait que ce processus soit un mouvement et non un champs stochastique rend encore l'application plus difficile, qui plus est en géostatistique. Ces constatations me permettent du coup de proposer assez logiquement une piste possible pour continuer ces recherches. Complicquer la modélisation pour obtenir un drap stable fractionnaire. Pour ce faire, il convient d'appliquer à un drap stable un filtre à l'aide d'une fonction noyau. D'une part, cela permet de gommer les effets indésirables associés aux axes. D'autre part, cela permet d'introduire de la dépendance spatiale, à longue étendue même. Le sujet est déjà étudié, mais beaucoup reste à faire, en particulier dans l'estimation des paramètres. Fréquemment, le seul paramètre estimé étant celui associé à la fonction noyau, directement lié à la dépendance. Quant aux applications, elles existent, c'est certain, tant dans le traitement de données géologiques que dans l'imagerie spatiale.

Ceci étant dit, il ne reste plus qu'à mettre le véritable point final à ce travail et je m'exécute aussitôt.

Bibliographie

- Adler, R. J. (1981). *The Geometry of Random Fields*. John Wiley & Sons.
- Adler, R. J., Monrad, D., Scissors, R. H., and Wilson, R. (1983). Representations, decompositions and sample function continuity of random fields with independent increments. *Stochastic Processes and their Applications*, **15**, 3–30.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B. Methodological*, **36**, 192–236.
- Ceroli, A. and Riani, M. (1999). The ordering of spatial data and the detection of multiple outliers. *Journal of Computational and Graphical Statistics*, **8**, 239–258.
- Chambers, J. M., Mallows, C. L., and Stuck, B. W. (1976). A method for simulating stable random variables. *Journal of the American Statistical Association*, **71**, 340–344.
- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829–836.
- Cox, D. R. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, **91**, 729–737.
- Cressie, N. (1986). Kriging nonstationary data. *Journal of the American Statistical Association*, **81**, 625–634.
- Cressie, N. (1989). Ergodicity for time series and spatial processes. *Journal of Statistical Computation and Simulation*, **32**, 61–63.
- Cressie, N. (1990). The origins of kriging. *Mathematical Geology*, **22**, 239–252.
- Cressie, N. and Hawkins, D. M. (1980a). Robust estimation of the variogram. I. *Journal of the International Association for Mathematical Geology*, **12**, 115–125.

- Cressie, N. A. C. (1985). Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology*, **17**, 563–586.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*. John Wiley & Sons Inc., New York, revised reprint.
- Cressie, N. A. C. and Hawkins, D. M. (1980b). Robust estimation of the variogram. I. *Journal of the International Association for Mathematical Geology*, **12**, 115–125.
- Dalang, R. C. (2003). Level sets and excursions of the Brownian sheet. In : *Topics in spatial stochastic processes (Martina Franca, 2001)*, volume 1802 of *Lecture Notes in Math.*, 167–208. Springer, Berlin.
- Dalang, R. C. and Khoshnevisan, D. (2004). Recurrent lines in two-parameter isotropic stable Lévy sheets. *Stochastic Processes and their Applications*, **114**, 81–107.
- Dalang, R. C. and Walsh, J. B. (1992). The sharp Markov property of Lévy sheets. *The Annals of Probability*, **20**, 591–626.
- DuMouchel, W. H. (1973). On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *The Annals of Statistics*, **1**, 948–957.
- Fama, E. F. and Roll, R. (1971). Parameter estimates for symmetric stable distributions (Ref : V63 p817-36). *Journal of the American Statistical Association*, **66**, 331–338.
- Fang, K. T., Kotz, S., and Ng, K. W. (1990). *Symmetric multivariate and related distributions*, volume 36 of *Monographs on Statistics and Applied Probability*. Chapman and Hall Ltd., London.
- Feller, W. (1971). *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons Inc., New York.
- Fournier, B. and Furrer, R. (2005a). Automatic mapping in the presence of substitutive errors : A robust kriging approach. *Applied GIS*, **1(2)**, doi :10.2104/ag050012.
- Fournier, B. and Furrer, R. (2005b). Automatic mapping in the presence of substitutive errors : A robust kriging approach. *Office for Official Publications of the European Communities, Luxembourg, EUR*, **21595 EN**, .
- Genton, M. G. (1998a). Highly robust variogram estimation. *Mathematical Geology*, **30**, 213–221.
- Genton, M. G. (1998b). Variogram fitting by generalized least squares using an explicit formula for the covariance structure. *Mathematical Geology*, **30**, 323–345.

- Genton, M. G. (2000). The correlation structure of Matheron's classical variogram estimator under elliptically contoured distributions. *Mathematical Geology*, **32**, 127–137.
- Genton, M. G. (2001). Robustness problems in the analysis of spatial data. In : *Spatial statistics : methodological aspects and applications*, volume 159 of *Lecture Notes in Statist.*, 21–37. Springer, New York.
- Hampel, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.*, **42**, 1887–1896.
- Hawkins, D. M. and Cressie, N. (1980). Robust kriging—a proposal. *Journal of the International Association for Mathematical Geology*, **16**, 3–18.
- Heagerty, P. J. and Lele, S. R. (1998). A composite likelihood approach to binary spatial data. *Journal of the American Statistical Association*, **93**, 1099–1111.
- Huber, P. J. (1981). *Robust Statistics*. John Wiley & Sons Inc., New York.
- Kogon, S. M. and Williams, D. B. (1998). Characteristic function based estimation of stable distribution parameters. In : *A practical guide to heavy tails : statistical techniques and applications*, 311–335. Birkhauser Boston Inc., Cambridge, MA, USA.
- Koutrouvelis, I. A. (1980). Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, **75**, 918–928.
- Koutrouvelis, I. A. (1981). An iterative procedure for the estimation of the parameters of stable laws. *Communications in Statistics. B. Simulation and Computation*, **10**, 17–28.
- Kuk, A. Y. and Nott, D. J. (2000). A pairwise likelihood approach to analysing correlated binary data. *Statistics & Probability Letters*, **47**, 329–335.
- Lévy, P. (1924). Théorie des erreurs. La loi de Gauss et les lois exceptionnelles. *Bulletin de la Société Mathématique de France*, **52**, 49–85.
- Matérn, B. (1986). *Spatial Variation*. Springer-Verlag, Berlin, second edition.
- Matheron, G. (1962). Traité de géostatistique appliquée, tome 1. *Mémoires du Bureau de Recherches Géologiques et Minières*, **24**, .
- McCulloch, J. H. (1986). Simple consistent estimators of stable distribution parameters. *Communications in Statistics. B. Simulation and Computation*, **15**, 1109–1136.
- Mercer, W. and Hall, A. (1911). The experimental error of fields trials. *Journal of Agricultural Science (Cambridge)*, **4**, 107–132.

- Mittnik, S., Rachev, S. T., Doganoglu, T., and Chenyao, D. (1999). Maximum likelihood estimation of stable Paretian models. *Mathematical and Computer Modelling*, **29**, 275–293.
- Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. John Wiley & Sons Inc., New York.
- Nolan, J. P. (1998). Parameterizations and modes of stable distributions. *Statistics & Probability Letters*, **38**, 187–195.
- Nolan, J. P. (2001). Maximum likelihood estimation and diagnostics for stable distributions. In : *Lévy processes*, 379–400. Birkhäuser Boston, Boston, MA.
- Nott, D. J. and Rydén, T. (1999). Pairwise likelihood methods for inference in image models. *Biometrika*, **86**, 661–676.
- Painter, S. (1995). Random fractal models of heterogeneity : The Lévy-stable approach. *Mathematical Geology*, **27**, 813–830.
- Press, S. J. (1972). Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association*, **67**, 842–846.
- Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, **88**, 1273–1283.
- Schabenberger, O. and Gotway, C. A. (2005). *Statistical methods for spatial data analysis*. Texts in Statistical Science Series. Chapman & Hall/CRC, Boca Raton, FL.
- Srivastava, R. M. (1986). Philip and Watson – quo vadunt ? *Mathematical Geology*, **18**, 141–146.
- Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In : *Contributions to probability and statistics*, 448–485. Stanford Univ. Press, Stanford, Calif.
- Yamazato, M. (1978). Unimodality of infinitely divisible distribution functions of class L . *The Annals of Probability*, **6**, 523–531.
- Zimmerman, D. L. and Zimmerman, M. B. (1991). A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictors. *Technometrics*, **33**, 77–91.
- Zolotarev, V. M. (1986). *One-dimensional stable distributions*, volume 65 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI.

Curriculum Vitae

Baptiste Fournier

Contact

Baptiste Fournier
Avenue du Temple 5
1012 Lausanne
baptiste.fournier@a3.epfl.ch

Date de naissance : 12 novembre 1976
Origine : Nendaz
Nationalité : Suisse

Education

Depuis septembre 2002	Etudiant–doctorant et assistant au sein de la chaire de statistique appliquée de l’EPFL
1996–2002	Etudes de mathématiques à l’Institut de Mathématiques de l’EPFL : obtention du diplôme d’ <i>Ingénieur mathématicien diplômé EPFL</i> . Travail de diplôme effectué au Centre National de Recherche Atmosphérique, Colorado, USA.
1991–1996	Obtention d’une <i>Maturité cantonale, Type C</i> au Lycée–Collège des Creusets, Sion.

Expériences professionnelles

Depuis 2005	Remplacement du Prof. S. Morgenthaler pour des cours de statistique pour étudiant en ingénierie et en mathématique.
2004	Participation au Concours International d’Interpolation Spatiale (SIC’04).
2003	Organisateur de la 13 ^{ème} Conférence Européenne des Jeunes Statisticiens (EYSM’03) à Ovronnaz, Suisse.
Depuis 2002	Assistant pour différents cours de statistique appliquée à l’EPFL.
2001	Conseil et analyse statistique pour l’entreprise METEO-RISK, Sion.

Autre activité professionnelle

Brevet fédéral de guide de haute montagne (UIAGM)

